

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Méthodes de Monte-Carlo à chaînes de Markov pour l'analyse de données cosmologiques

SPIES, Vincent

Award date:
2014

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



UNIVERSITE DE NAMUR

Faculté des Sciences

**MÉTHODES DE MONTE-CARLO À CHAÎNES DE MARKOV
POUR L'ANALYSE DE DONNÉES COSMOLOGIQUES**

**Mémoire présenté pour l'obtention
du grade académique de master en « Sciences Mathématiques »**

Vincent SPIES

Août 2014



UNIVERSITE DE NAMUR

Faculté des Sciences

**MÉTHODES MONTE-CARLO À CHAÎNES DE MARKOV
POUR L'ANALYSE DE DONNÉES COSMOLOGIQUES**

Promoteur : André Füzfa

Maître de stage : Christophe Ringeval

**Mémoire présenté pour l'obtention
du grade académique de master en « Sciences Mathématiques »**

Vincent Spies

Août 2014

Résumé

Les méthodes Monte-Carlo à chaînes de Markov sont utilisées dans de nombreux domaines, elle permettent de générer une séquence d'échantillons suivant une certaine densité de probabilité. Cela est fort utile en inférence bayésienne, qui consiste à estimer les paramètres d'un modèle à partir des observations d'un système physique étudié. L'inférence bayésienne pose le problème sous forme de fonction de densité a posteriori. Toute l'information sur les paramètres du modèle est contenue dans cette densité. Grâce aux méthodes MCMC, il est possible, quelque soit la fonction, de générer des échantillons de ces paramètres.

Nous l'appliquons pour estimer les paramètres cosmologiques, en comparant les données observées de supernovae avec le calcul de la distance lumineuse, dépendant de ces paramètres cosmologiques.

Abstract

Markov chain Monte Carlo are used in several fields to sampler from a given probability distribution. This is useful in bayesian inference, which consists in estimating model parameters from observations of a given physical system. Problem is formed as a *a posteriori* probability distribution by inference bayesian. All information on parameters of the model is contained by this distribution. With MCMC methods, it is possible, for any probability distribution, to make sampling from these parameters.

We use this method to estimate cosmological parameters, by comparing observed data of supernova with luminosity distance computing, which depends on these cosmological parameters.

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de mes études.

Un grand merci à Monsieur Füzfa, qui m'a bien conseillé tout au long de l'année, et sans qui ce travail n'aurait pu être réalisé.

J'adresse aussi mes remerciements à Monsieur Ringeval qui m'a guidé lors de mon stage au CP3 à Louvain-La-Neuve. Merci pour le temps que vous m'avez consacré.

Merci également à Monsieur Clesse, qui a toujours répondu à mes questions en cosmologie.

Et un dernier merci pour ma famille, pour sa contribution, pour m'avoir soutenu particulièrement durant ces derniers mois.

Table des matières

Introduction	1
1 Notions de probabilité	3
1.1 Probabilité	4
1.2 Variables aléatoires réelles	4
1.3 Vecteurs aléatoires	6
1.4 Espérance et variance	7
1.5 Indépendance de variables aléatoires	8
1.6 Probabilité conditionnelle	9
2 Inférence bayésienne	11
2.1 Conceptualisation de l'étude d'un système physique	12
2.2 Principe de l'inférence bayésienne	13
2.3 Maximum de vraisemblance pour des erreurs gaussiennes . . .	14
2.3.1 Modèle d'inversion déterministe	16
2.3.2 Modèle d'inversion probabiliste	17
3 Méthodes MCMC	21
3.1 Introduction	22
3.2 Méthode de l'inverse	22
3.3 Méthode du rejet	24
3.4 Méthode de Monte-Carlo par chaînes de Markov	26
3.4.1 Chaînes de Markov	27
3.4.2 Algorithme de Metropolis-Hastings	33

4 Applications introductives sur l'inférence bayésienne	41
4.1 Outils	42
4.2 Régression linéaire	42
4.2.1 Données du problème	43
4.2.2 Résolution	43
4.3 Épicentre d'un événement sismique	45
4.3.1 Moment de l'événement sismique connu	46
4.3.2 Moment de l'événement sismique inconnu	48
5 Existence de l'énergie noire	51
5.1 Modélisation	52
5.2 Univers plat	53
5.3 Univers sphérique, plat ou hyperbolique	56
6 Conclusions et perspectives	59
Bibliographie	61
Annexe	63
A Calcul de la convergence des suites	63

Introduction

L'origine de ce mémoire est de présenter la méthode de l'inférence bayésienne et de l'utiliser pour déterminer les paramètres cosmologiques en fonction de données observées.

Lorsque l'on souhaite analyser un système physique, on crée généralement un modèle dont on ajuste au mieux les paramètres afin de pouvoir prédire les valeurs des données du système. L'inversion bayésienne va consister à observer le système afin de pouvoir recueillir des informations sur les paramètres du modèle, en tenant compte de la connaissance a priori que l'on peut avoir sur ces paramètres.

À l'inverse d'une méthode dite déterministe, l'information recueillie sur les paramètres du modèle par inférence bayésienne sera sous forme d'une densité de probabilité : nous pourrons construire des intervalles de confiance sur les valeurs probables des paramètres du modèle en fonction des données observées. Nous verrons que ces paramètres et ces observations seront considérées comme des variables aléatoires.

Nous commencerons par rappeler les notions de probabilité nécessaires à la présentation de l'inférence bayésienne, cette dernière se basant sur le théorème de Bayes.

Ensuite nous détaillerons théoriquement le principe de l'inférence bayésienne dans le cas de l'étude d'un système physique quelconque. Nous supposons que les erreurs des données observées suivent une loi gaussienne, et détaille-

rons analytiquement les résultats dont nous aurons besoin dans la résolution d'un problème concret.

Comme l'inférence bayésienne donne des informations sur les paramètres du modèle sous forme d'une fonction de densité, il nous faut introduire des méthodes qui permettent d'engendrer des échantillons suivant ces densités de probabilité. Nous présenterons les méthodes de l'inverse et du rejet, qui sont utiles lorsque l'on a une fonction de densité sous une forme analytique et inversible, ou à une seule variable. Dans le cas contraire, ces méthodes sont moins efficaces (voire non applicables). Nous détaillerons alors les méthodes de Monte-Carlo par chaînes de Markov qui sont plus adaptées à la résolution de ces problèmes.

Enfin, une fois tous les requis théoriques présentés, nous appliquerons l'inférence bayésienne à trois exemples concrets : un problème de régression linéaire simple, la détection de l'origine d'un événement sismique via les données que reçoivent les stations environnantes, et la détermination des paramètres cosmologiques sur base des données observées des supernovae.

Chapitre 1

Notions de probabilité

Dans ce chapitre nous introduirons les concepts de probabilité qui seront nécessaires au développement de la théorie de l'inférence bayésienne et des chaînes de Markov.

Nous définirons ce qu'est une probabilité, une variable aléatoire réelle dans le cas continu et discret, et une fonction de densité. Nous généralisons ensuite ces concepts à plusieurs dimensions via les vecteurs aléatoires. Enfin, nous citerons le théorème de Bayes et introduirons densités conditionnelles.

1.1 Probabilité

Nous commençons par définir ce qu'est une probabilité en nous basant sur la théorie de la mesure (une probabilité est une mesure à valeurs dans $[0, 1]$). Notons que cette partie se base sur le cours de probabilité de A. Hardy [1].

Définition 1.1 *Probabilité*

Soit Ω un ensemble quelconque et \mathcal{A} une tribu sur Ω . La fonction

$$P : \Omega \longrightarrow \mathbb{R}$$

est une probabilité ssi toutes les conditions suivantes sont satisfaites :

1. $P(\Omega) = 1$
2. $P(\emptyset) = 0$
3. $\forall A \in \mathcal{A} : P(A) \geq 0$
4. $\forall (A_i)_{i \in \mathbb{N}}$ tq $\forall i \neq j : A_i \cap A_j = \emptyset$, on a : $P[\bigcup_{i \in \mathbb{N}} A_i] = \sum_{i \in \mathbb{N}} P(A_i)$

1.2 Variables aléatoires réelles

Le but des méthodes Monte-Carlo à chaînes de Markov étant d'engendrer des échantillons (représentés par une variable aléatoire) suivant une certaine fonction de densité, nous donnons la définition de ces deux concepts. Nous détaillons également la différence entre variables continues et discrètes, étant donné que nous serons amenés à engendrer l'une ou l'autre en fonction du système à analyser.

Définition 1.2 *Variable aléatoire réelle*

Une variable aléatoire réelle est une fonction $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ mesurable, où $\mathcal{B}_{\mathbb{R}}$ est la tribu borélienne.

Définition 1.3 *Loi de probabilité d'une variable aléatoire réelle*

Soient (Ω, \mathcal{A}, P) un espace probabilisé et $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire réelle, alors la fonction P_X définie par

$$P_X : \mathcal{B}_{\mathbb{R}} \longrightarrow \mathbb{R} : B \longmapsto P_X(B) = P(X^{-1}(B))$$

est une probabilité et est appelée loi de probabilité de la VAR X .

Remarque 1.1 On utilisera les notations suivantes :

- $P_X(B) = P(X^{-1}(B)) = P(X \in B)$
- $P_X([a, b]) = P(X \in [a, b]) = P(a \leq X \leq b)$

Définition 1.4 Fonction de répartition

La fonction de répartition F_X d'une variable aléatoire X est définie par :

$$F_X : \mathbb{R} \longrightarrow \mathbb{R} : x \longmapsto F_X(x) = P(X \leq x)$$

Définition 1.5 Variables aléatoires discrètes et continues

Considérons une variable aléatoire X et sa fonction de répartition F_X , et notons D_F l'ensemble des points de discontinuité de F_X défini par

$$D_F = \{x \in \mathbb{R} \mid F_X(x) - F_X(x^-) > 0\}$$

On dira alors :

- X est une variable aléatoire discrète si $P(X \in D_F) = 1$, et
- X est une variable aléatoire continue si $D_F = \emptyset$

Remarque 1.2 Notons qu'une variable aléatoire discrète prend un nombre fini ou infini dénombrable de valeurs, tandis qu'une variable aléatoire continue prend un nombre infini non dénombrable de valeurs.

Définition 1.6 Fonction de densité d'une variable aléatoire réelle

Soit X une variable aléatoire réelle.

1. **Cas discret** : Si X est une variable aléatoire discrète, sa loi de probabilité peut s'écrire comme :

$$P_X = \sum_{x \in D_F} P(X = x) \delta_x$$

où δ_x est la masse de Dirac définie par :

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

On notera $p_X(x) = P(X = x)$

2. **Cas continu** : Si X est une variable aléatoire continue, sa fonction de densité p_X est telle que, $\forall A \in \mathcal{B}_{\mathbb{R}}$:

$$P_X(A) = \int_A p_X(x) dx$$

1.3 Vecteurs aléatoires

Les définitions de la section précédente se généralisent aux variables aléatoires à valeurs dans \mathbb{R}^n . Nous en aurons besoin, car un système nécessitant la détermination de plusieurs paramètres requiert l'utilisation de fonctions de densité à plusieurs variables.

Nous calculerons, au chapitre suivant consacré à l'inférence bayésienne, la loi jointe de variables aléatoires.

Dans le cas où une chaîne de Markov converge vers une fonction de densité à plusieurs variables, cette chaîne est en réalité une suite de vecteurs aléatoires, où chacune des variables convergent vers la fonction de densité marginale associée. Un exemple sera réalisé au chapitre 3.

Définition 1.7 Vecteur aléatoire

Un vecteur aléatoire X à valeur dans \mathbb{R}^n est un n -uplet (X_1, \dots, X_n) composé de variables aléatoires réelles à valeur dans \mathbb{R} .

Définition 1.8 Loi de probabilité d'un vecteur aléatoire

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. La loi P_X de X (ou loi jointe des variables aléatoires X_1, \dots, X_n) est définie par

$$\begin{aligned} P_X : \quad \mathcal{B}_{\mathbb{R}^n} &\longrightarrow \mathbb{R} \\ B = B_1 \times \dots \times B_n &\longmapsto P_X(B) = P(X^{-1}(B_1 \times \dots \times B_n)) \end{aligned}$$

Remarque 1.3 Nous utiliserons la notation suivante pour la loi d'un vecteur

aléatoire X :

$$\begin{aligned} P_X(B) &= P(X^{-1}(B_1 \times \cdots \times B_n)) \\ &= P((X_1, \dots, X_n) \in B_1 \times \cdots \times B_n) \\ &= P(X_1 \in B_1, \dots, X_n \in B_n) \end{aligned}$$

Définition 1.9 Fonction de répartition d'un vecteur aléatoire

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. La fonction F_X de répartition de X est définie par :

$$\begin{aligned} F_X : \quad \mathbb{R}^n &\longrightarrow [0, 1] \\ (x_1, \dots, x_n) &\longmapsto F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

Définition 1.10 Fonction de densité d'un vecteur aléatoire

La fonction de densité p_X d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ (ou fonction de densité conjointe de (X_1, \dots, X_n)) est la fonction p_X définie presque partout telle que $\forall B \in \mathcal{B}_{\mathbb{R}^n}$:

$$P_{X_1, \dots, X_n}(B_1 \times \cdots \times B_n) = \int \cdots \int_{B_1 \times \cdots \times B_n} p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

Définition 1.11 Fonction de densité marginale

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. La fonction de densité marginale $p_{X_i}(x_i)$ est définie comme :

$$\begin{aligned} p_{X_i}(x_i) &= \int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dt_{i-1} \int_{-\infty}^{+\infty} dt_{i+1} \int_{-\infty}^{+\infty} \\ &\quad \cdots \int_{-\infty}^{+\infty} p_X(t_1, \dots, t_{i-1}, x_i, t_{i+1}, \dots, t_n) dt_n \end{aligned} \quad (1.1)$$

1.4 Espérance et variance

L'espérance et la variance sont des concepts fondamentaux dans l'étude de variables aléatoires. Néanmoins, ceux-ci nécessitent un calcul d'intégration. Nous verrons qu'il est possible, grâce aux méthodes de Monte-Carlo, de résoudre numériquement ces intégrales.

Définition 1.12 *Espérance d'une variable aléatoire*

Soit $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire, on définit son espérance $E[X]$ comme :

$$E[X] = \int_{\mathbb{R}} x p_X(x) dx$$

Remarque 1.4 Si X est une variable aléatoire discrète, alors l'espérance vaut :

$$E[X] = \sum_{x \in X(\Omega)} x P(X = x)$$

Définition 1.13 *Variance d'une variable aléatoire*

Soit $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire, on définit sa variance $V[X]$ comme :

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= \int_{\mathbb{R}} (x - E[X])^2 p_X(x) dx \end{aligned}$$

1.5 Indépendance de variables aléatoires

L'indépendance entre variables aléatoires est une hypothèse qui reviendra souvent dans la suite du mémoire. Généralement, lorsqu'on étudie un système physique, on sait si les données observées sont dépendantes ou non. Cette hypothèse d'indépendance ne nécessite donc pas une vérification a posteriori.

Définition 1.14 *Indépendance de deux variables aléatoires réelles*

Soit X et Y deux variables aléatoires réelles. On dit que X et Y sont indépendantes (que l'on notera $X \sqcup Y$) ssi :

$$p_X(x) p_Y(y) = p_{X,Y}(x, y) \quad \forall x, y \in \mathbb{R}$$

où p_X et p_Y sont respectivement les lois de densité marginale par rapport à X et à Y de (X, Y) , et $p_{X,Y}$ est la loi de densité conjointe de (X, Y) .

Proposition 1.1 *Densité d'une somme de variables aléatoires indépendantes*

Soient X et Y deux variables aléatoires indépendantes. Alors la fonction de densité de p_{X+Y} est le produit de convolution entre p_X et p_Y , i.e., $\forall z \in \mathbb{R}$:

$$\begin{aligned} p_{X+Y}(z) &= (p_X * p_Y)(z) \\ &= \int_{-\infty}^{+\infty} p_X(z-t)p_Y(t)dt \\ &= \int_{-\infty}^{+\infty} p_Y(z-t)p_X(t)dt \\ &= p_{Y+X}(z) \end{aligned}$$

1.6 Probabilité conditionnelle

Nous introduisons ici le théorème de Bayes, fondement de l'inférence bayésienne. Celui-ci se base sur les fonctions de densités conditionnelles, définies par l'intermédiaire de densités conjointes. Nous pouvons constater que l'indépendance entre variables aléatoires simplifiera ces définitions.

Définition 1.15 *Fonction de densité conditionnelle*

Soient X et Y deux variables aléatoires réelles. La densité de X sachant que $Y = y_0$, notée $p_{X|Y}(x|y_0)$, est définie comme suit [17] :

$$p_{X|Y}(x|y_0) = \frac{p_{X,Y}(x, y_0)}{p_Y(y_0)}$$

où $p_{X,Y}$ est la densité conjointe de (X, Y) .

Théorème 1.1 *Théorème de Bayes*

L'énoncé de ce théorème provient de la section 2.1.3 de [10].

Soient X et Y deux variables aléatoires réelles. Par la définition de loi de probabilité conjointe, on a, $\forall x, y \in \mathbb{R}$:

$$p_{X,Y}(x, y) = p_{Y,X}(y, x)$$

Donc, par la définition de probabilité conditionnelle, on a les égalités :

$$p_{X,Y}(x, y) = p_{X|Y}(x|y) p_Y(y) = p_{Y|X}(y|x) p_X(x) = p_{Y,X}(y, x)$$

D'où :

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y) p_Y(y)}{p_X(x)}$$

De plus, par la définition de la loi de densité marginale, on a :

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{+\infty} p_{X,Y}(x,y) dy \\ &= \int_{-\infty}^{+\infty} p_{X|Y}(x|y) p_Y(y) dy \end{aligned}$$

On obtient la formule du théorème de Bayes :

$p_{Y X}(y x) = \frac{p_{X Y}(x y) p_Y(y)}{\int_{-\infty}^{+\infty} p_{X Y}(x y) p_Y(y) dy} \quad (1.2)$
--

Nous avons à présent défini tous les concepts dont nous avons besoin pour décrire la méthode de l'inférence bayésienne. Dans le prochain chapitre, nous détaillerons cette méthode et monterons où intervient le théorème de Bayes.

Chapitre 2

Inférence bayésienne

Dans cette partie nous décrirons la théorie entourant la méthode d'inférence bayésienne. Tout d'abord, nous conceptualiserons de manière générale l'étude d'un système physique, puis nous introduirons la fonction de vraisemblance. Nous montrerons son lien avec le théorème de Bayes, ce qui nous conduira à la méthode de l'inférence bayésienne.

Comme dit lors de l'introduction, nous considérerons des erreurs d'observations gaussiennes. Nous construirons, dès lors, la fonction de vraisemblance pour ces erreurs, selon deux considérations conceptuelles différentes que nous qualifierons de déterministe et probabiliste.

2.1 Conceptualisation de l'étude d'un système physique

Pour étudier un système physique, nous avons généralement des données observées, mesurées avec ou sans erreurs, et un modèle qui nous permet de déduire les observations. Sur base de la section C. Koot [10], nous détaillons ceci. Soient :

- $\theta = (\theta_1, \dots, \theta_n)$: les paramètres d'un système dont on cherche à connaître la fonction de densité.
- $M(\theta)$: le modèle destiné à prédire les données du système (les paramètres du système sont également considérés comme paramètres du modèle).
- $\pi(\theta)$: la fonction de densité a priori des paramètres du modèle. Celle-ci permet de prendre en considération les connaissances a priori sur les paramètres.
- $z = (z_1, \dots, z_n)$: les prédictions des données du système par le modèle.
- $d = (d_1, \dots, d_m)$: les observables du système, qui diffèrent généralement des prédictions z en raison d'erreurs de mesures.
- $e = (e_1, \dots, e_n)$: les erreurs de mesures des données observées du système. Il est possible que les données soient mesurées, par exemple, avec une erreur suivant une certaine densité de probabilité avec un écart type donné. Ce dernier serait alors un bon indicateur des erreurs de mesures.

Notre objectif est de déterminer les paramètres tels que le modèle prédise, au mieux, les observables. Une méthode habituelle¹ est de construire la *fonction du maximum de vraisemblance*, qui est une fonction des paramètres du modèle, que l'on doit maximiser afin de réduire l'erreur entre les données observées et celles prédites par le modèle. Celle-ci est définie à la section suivante.

1. Selon « Lecture 5 » de P. Protopapas [13]

2.2 Principe de l'inférence bayésienne

Dans cette partie, nous définissons la fonction du maximum de vraisemblance, montrons son lien avec le théorème de Bayes, et déterminons son expression analytique pour des erreurs de mesures gaussiennes.

Il est important de noter que nous ferons, dorénavant, un abus de langage fréquent dans la théorie de l'inférence bayésienne : il s'agit de représenter identiquement les variables aléatoires et leurs réalisations quand le contexte le permet (i.e. quand il n'y a pas d'ambiguïté). Dans le cas contraire, nous noterons la variable aléatoire par une lettre majuscule, et les valeurs prises par celle-ci par la lettre minuscule correspondante.

Nous définissons la fonction du maximum de vraisemblance :

Définition 2.1 *Fonction du maximum de vraisemblance*

Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées, et $\theta \in \Theta$ un paramètre connu.

Alors la fonction de vraisemblance $L(x_1, \dots, x_n | \theta) = L(\theta)$ s'exprime comme :

$$L(\theta) = \prod_{i=1}^n p_{X_i}(x_i | \theta)$$

Nous pouvons montrer que la connaissance de cette fonction (sous n'importe quelle forme, pas forcément analytique, mais que l'on sache au moins l'évaluer) permet d'inférer sur les paramètres θ de manière à prédire au mieux les données du système.

Pour ce faire, on utilise ce que l'on va définir comme étant la fonction a posteriori : $p(\theta | d)$. Celle-ci est une fonction de densité conditionnelle, donnant toutes les informations sur les paramètres θ recherchés en fonction des mesures d . Si on connaît cette densité a posteriori, il suffira d'engendrer des échantillons de cette dernière, on aura alors un ensemble de paramètres θ décrivant le système avec une probabilité $p(\theta | d)$. C'est donc bien ce que l'on

recherche. Nous pouvons lier cette probabilité à la fonction de vraisemblance grâce au théorème de Bayes par l'égalité suivante :

$$p(\theta|d) = \frac{L(d|\theta)\pi(\theta)}{\int_{\Theta} L(d|\theta)\pi(\theta)d\theta}. \quad (2.1)$$

Nous verrons au chapitre 3 que pour engendrer une variable aléatoire suivant une certaine fonction cible, il est suffisant de connaître celle-ci à un facteur près. L'équation (2.1) peut donc se réécrire comme suit :

$$p(\theta|d) \propto L(d|\theta)\pi(\theta).$$

Note : Nous appellerons également « samples » les réalisations d'une variable aléatoire engendrée par une densité de probabilité cible.

La fonction de densité a posteriori $p(\theta|d)$ est donc bien déterminée si $L(d|\theta)$ est connue (on appelle aussi cette fonction la « likelihood »). Nous déterminons une expression analytique de celle-ci à la section suivante, dans le cas où les erreurs suivent une loi gaussienne (pour rappel, nous ferons toujours cette hypothèse dans ce mémoire).

2.3 Maximum de vraisemblance pour des erreurs gaussiennes

Nous allons déterminer une expression analytique de la likelihood. Pour cela, faisons l'hypothèse que les données observées d peuvent être écrites comme la somme entre les données prédites par le modèle, et les erreurs de mesures ; i.e. :

$$d = z + e.$$

Nous noterons $p_E(e)$ la densité de probabilité des erreurs de mesures et $m(z|\theta)$ la probabilité conditionnelle de z en fonction de θ , i.e. $m(z|\theta) = p_Z(z|\theta)$.

Nous pouvons exprimer $L(d|\theta)$ comme suit :

$$\begin{aligned} L(d|\theta) &= p_D(d|\theta) \\ &= p_{Z+E}(d|\theta) \\ &= \int_{-\infty}^{+\infty} p_E(d-z|\theta) m(z|\theta) dz, \end{aligned} \quad (2.2)$$

où l'on a fait l'hypothèse d'indépendance entre Z et E pour la dernière égalité.

Les densités marginales $L(d_i|\theta)$ valent donc :

$$\begin{aligned} L(d_i|\theta) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} m(z|\theta) p_E(e-z) d(d_1) \cdots d(d_{i-1}) d(d_{i+1}) \cdots d(d_m) dz \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} m(z|\theta) \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_E(d-z) d(d_1) \cdots d(d_{i-1}) d(d_{i+1}) \cdots d(d_m) \right] dz \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} m(z|\theta) p_{E_i}(d_i - z_i) dz \end{aligned} \quad (2.3)$$

$$\begin{aligned} &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} m(z|\theta) dz_1 \cdots dz_{i-1} dz_{i+1} \cdots dz_n \right] p_{E_i}(d_i - z_i) dz_i \\ &= \int_{-\infty}^{+\infty} m(z_i|\theta) p_{E_i}(d_i - z_i) dz_i, \end{aligned} \quad (2.4)$$

où l'on a utilisé en (2.3) et en (2.4) la définition de loi de densité marginale.

De plus, par l'hypothèse d'indépendances entre les données mesurées, nous pouvons exprimer la fonction de vraisemblance par :

$$L(d|\theta) = \prod_{i=1}^m L(d_i|\theta).$$

Avant de déterminer l'expression de $L(d|\theta)$, nous opposons deux conceptualisations. La première suggère un modèle d'inversion dit déterministe : nous supposons que le modèle $M(\theta)$ prédit exactement les données du système, i.e. on suppose qu'il n'y a pas d'erreurs sur les paramètres. Notons que nous considérons toujours des erreurs de mesures sur les données observées, mais nous supposons, par exemple, que celles-ci ont plus d'importance que les erreurs sur les paramètres (qui sont, comme précisé par C. Koot [10], souvent liées à la méconnaissance du système). À l'inverse, dans le cas d'un modèle

d'inversion probabiliste, celui-ci ne prédit pas une valeur pour les données du système, mais une distribution de valeurs (le modèle est donc sous la forme d'une densité de probabilité). Dans les deux sections qui suivent, nous calculons la likelihood selon chacune des conceptualisations.

2.3.1 Modèle d'inversion déterministe

Dans le cas d'un modèle d'inversion déterministe, on suppose que le modèle prédit exactement les données du système. On a donc :

$$M(\theta) = z. \quad (2.5)$$

Nous pouvons (par [10]) dès lors considérer la fonction de densité $m(z|\theta)$ comme valant 1 si $M(\theta) = z$ et 0 sinon. Ceci peut s'exprimer grâce à la *distribution de dirac* que nous définissons, sur base de O. Garet [7] et S. Chakraborty [4], comme suit :

Définition 2.2 *Fonction de Dirac*

La distribution de Dirac $\delta(x)$ est une fonction à valeur réelle telle que :

1. $\int_{-\infty}^{+\infty} \delta(x)dx = 1,$
2. $\int_a^b f(x)\delta(x - x_0)dx = f(x_0), \forall x_0 : a < x_0 < b.$

La fonction $m(z|\theta)$ valant 1 en $z = M(\theta)$ et 0 sinon, on a donc l'égalité :

$$m(z|\theta) = \delta(z - M(\theta)).$$

Nous pouvons exprimer la relation (2.2) en fonction de la distribution de Dirac et utiliser les propriétés de cette dernière ; on obtient :

$$\begin{aligned} \int_{-\infty}^{+\infty} p_E(d - z|\theta)m(z|\theta)dz &= \int_{-\infty}^{+\infty} p_E(d - z|\theta)\delta(z - M(\theta))dz \\ &= p_E(d - M(\theta)), \end{aligned} \quad (2.6)$$

où l'on a substitué z par $M(\theta)$ dans $p_E(d - z|\theta)$.

En supposant que les erreurs de mesures sont indépendantes, la fonction de

vraisemblance vaut finalement :

$$L(d|\theta) = \prod_{i=1}^m p_{E_i}(d_i - M_i(\theta)).$$

Donnons l'expression de $L(\theta)$ dans le cas où les erreurs suivent une loi gaussienne centrée en zéro et de variance σ^2 ; i.e. supposons que $E_i \sim N(0, \sigma^2)$, avec donc une fonction de densité pour E_i valant : $p_{E_i}(e_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{e_i}{\sigma}\right)^2\right]$, on obtient :

$$\begin{aligned} L(d|\theta) &= \prod_{i=1}^m p_{E_i}(d_i - M_i(\theta)) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^m \exp\left[-\frac{1}{2} \sum_{i=1}^m \left(\frac{d_i - M_i(\theta)}{\sigma}\right)^2\right] \end{aligned} \quad (2.7)$$

2.3.2 Modèle d'inversion probabiliste

Nous nous intéressons maintenant au cas du modèle d'inversion probabiliste, c'est-à-dire pour lequel les données du système ne sont pas prédites de manière exacte par le modèle, car des erreurs peuvent figurer dans ce dernier. En notant ϵ la variable aléatoire représentant les erreurs du modèle, nous pouvons établir dès lors la relation suivante :

$$z = M(\theta) + \epsilon.$$

Nous voyons la différence avec la relation de l'équation (2.5) : ici, $m(z|\theta)$ ne peut pas être représenté simplement par la distribution de Dirac. Néanmoins, z a la forme d'une somme de deux termes que nous considérons comme variables aléatoires indépendantes. Le modèle seul $M(\theta)$ étant toujours déterministe, nous pouvons effectuer le produit de convolution en utilisant la fonction de Dirac pour représenter $M(\theta)$. On obtient :

$$\begin{aligned} m(z|\theta) &= \int_{-\infty}^{+\infty} p_\epsilon(z - t) \delta(t - M(\theta)) dt \\ &= p_\epsilon(z - M(\theta)) \end{aligned}$$

La fonction de vraisemblance vaut donc :

$$L(d|\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_E(d - z) p_\epsilon(z - M(\theta)) dz.$$

À nouveau, supposons que les erreurs du modèle et les erreurs de mesures suivent une loi gaussienne, i.e. $\epsilon_i \sim N(0, \sigma_{M_i}^2)$ et $E_i \sim N(0, \sigma_i^2)$, alors :

$$m(z_i|\theta) = p_{\epsilon_i}(z_i - M_i(\theta)) = \frac{1}{\sigma_{M_i} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{z_i - M_i(\theta)}{\sigma_{M_i}} \right)^2 \right]$$

et

$$p_{E_i}(d_i - z_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{d_i - z_i}{\sigma_i} \right)^2 \right]$$

Si les mesures sont des variables aléatoires indépendantes, on a par l'égalité (2.4) :

$$\begin{aligned} L(d_i|\theta) &= \frac{1}{2\pi \sigma_{M_i} \sigma_i} \int_{-\infty}^{+\infty} \exp \left[-\frac{1}{2} \left(\frac{z_i - M_i(\theta)}{\sigma_{M_i}} \right)^2 \right] \\ &\quad \exp \left[-\frac{1}{2} \left(\frac{d_i - z_i}{\sigma_i} \right)^2 \right] dz_i \end{aligned} \quad (2.8)$$

Résolvons cette dernière intégrale de l'équation (2.8). Pour alléger les écritures, nous posons :

$$\begin{aligned} a_i &= z_i - M_i(\theta), \\ b_i &= d_i - M_i(\theta), \\ c_i &= \sqrt{\sigma_i^2 + \sigma_{M_i}^2} \end{aligned}$$

On obtient la suite des égalités suivantes :

$$\begin{aligned} L(d_i|\theta) &= \frac{1}{2\pi \sigma_{M_i} \sigma_i} \int_{\mathbb{R}} \exp \left[\frac{-a_i^2}{2\sigma_{M_i}^2} \right] \exp \left[\frac{-(d_i - z_i)^2}{2\sigma_i^2} \right] dz_i \\ &= \frac{1}{2\pi \sigma_{M_i} \sigma_i} \int_{\mathbb{R}} \exp \left[\frac{-a_i^2}{2\sigma_{M_i}^2} \right] \exp \left[\frac{-(b_i - a_i)^2}{2\sigma_i^2} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi\sigma_{M_i}\sigma_i} \int_{\mathbb{R}} \exp\left[\frac{-a_i^2}{2\sigma_{M_i}^2}\right] \exp\left[\frac{-1}{2\sigma_i^2} (b_i^2 + a_i^2 - 2b_i a_i)\right] dz_i \\
&= \frac{1}{2\pi\sigma_{M_i}\sigma_i} \exp\left[\frac{-b_i^2}{2\sigma_i^2}\right] \int_{\mathbb{R}} \exp\left[-\frac{c_i^2 a_i^2}{2\sigma_{M_i}^2 \sigma_i^2} + \frac{b_i a_i}{\sigma_i^2}\right] dz_i \\
&= \frac{1}{2\pi\sigma_{M_i}\sigma_i} \exp\left[\frac{-b_i^2}{2\sigma_i^2}\right] \int_{\mathbb{R}} \exp\left[-\frac{w^2}{2} + \frac{w b_i \sigma_{M_i}}{\sigma_i c_i}\right] \frac{\sigma_{M_i} \sigma_i}{c_i} dz_i \quad (2.9)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi} \exp\left[\frac{-b_i^2}{2\sigma_i^2}\right] \int_{\mathbb{R}} \frac{1}{c_i} \exp\left[-\frac{1}{2} \left(w - \frac{b_i \sigma_{M_i}}{\sigma_i c_i}\right)^2 + \frac{b_i^2 \sigma_{M_i}^2}{2\sigma_i^2 c_i^2}\right] dz_i \\
&= \frac{1}{2\pi c_i} \exp\left[-\frac{b_i^2}{2\sigma_i^2}\right] \exp\left[\frac{b_i^2 \sigma_{M_i}^2}{2\sigma_i^2 c_i^2}\right] \int_{\mathbb{R}} \exp\left[-\left(w - \frac{b_i \sigma_{M_i}}{\sigma_i c_i}\right)^2\right] dw \\
&= \frac{1}{2\pi c_i} \exp\left[-\frac{b_i^2}{2\sigma_i^2}\right] \exp\left[\frac{b_i^2 \sigma_{M_i}^2}{2\sigma_i^2 c_i^2}\right] \int_{\mathbb{R}} \exp\left[-\frac{1}{2} v^2\right] dv \quad (2.10)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi} c_i} \exp\left[\frac{b_i^2 (-c_i^2 + \sigma_{M_i}^2)}{2\sigma_i^2 c_i^2}\right] \quad (2.11) \\
&= \frac{1}{c_i} \exp\left[-\frac{b_i^2}{2c_i^2}\right].
\end{aligned}$$

REMARQUES :

- Pour l'égalité (2.9), on a fait le changement de variable :

$$w = \frac{c_i a_i}{\sigma_{M_i} \sigma_i}.$$

- Pour l'égalité (2.10), le changement de variable :

$$v = w - \frac{b_i \sigma_{M_i}}{\sigma_i c_i}.$$

- L'égalité (2.11) est obtenue en remarquant que

$$\int_{\mathbb{R}} e^{-\frac{1}{2} v^2} dv = \sqrt{2\pi}.$$

On obtient au final :

$$L(d_i|\theta) = \frac{1}{\sqrt{2\pi} c_i} \exp\left[-\frac{1}{2} \frac{b_i^2}{c_i^2}\right].$$

L'expression de la likelihood est donc la suivante :

$$\begin{aligned}
 L(d|\theta) &= \prod_{i=1}^m L(d_i|\theta) \\
 &= (2\pi)^{-\frac{m}{2}} \left(\prod_{i=1}^m \frac{1}{c_i} \right) \exp \left[- \sum_{i=1}^m \frac{1}{2} \frac{b_i}{c_i^2} \right] \\
 &= (2\pi)^{-\frac{m}{2}} \left(\prod_{i=1}^m \frac{1}{\sqrt{\sigma_i^2 + \sigma_{M_i}^2}} \right) \exp \left[- \sum_{i=1}^m \frac{1}{2} \frac{(d_i - M_i(\theta))^2}{\sigma_i^2 + \sigma_{M_i}^2} \right]
 \end{aligned}$$

On remarquera que le résultat final est proche de celui obtenu à l'équation (2.7), avec pour variance la somme des variances des erreurs de mesures et du modèle.

Maintenant que nous connaissons l'expression de la fonction a posteriori $p(\theta|d)$ pour des erreurs gaussiennes, nous allons, au prochain chapitre, détailler des méthodes qui permettent d'engendrer des variables aléatoires selon une densité de probabilité cible. Ces méthodes, appliquées à la fonction a posteriori, permettront de générer les paramètres du modèle.

Chapitre 3

Méthodes MCMC

Ce chapitre a pour objectif de construire une méthode de Monte-Carlo par chaîne de Markov, dans le but, notamment, d'engendrer des échantillons selon une fonction de densité cible donnée. Nous détaillons d'abord deux méthodes classiques qui fonctionnent pour des densités de probabilité particulières. Ensuite, afin d'avoir une méthode plus générale, nous introduirons l'algorithme de Metropolis Hastings, démontrerons sa convergence, et l'illustrerons par deux exemples. Cet algorithme étant basé sur les chaînes de Markov, nous leur consacrerons une section. Celle-ci définira ces chaînes sur base d'exemples.

3.1 Introduction

Nous avons vu que l'inférence bayésienne permettait de construire une fonction de densité a posteriori contenant toutes les informations sur les paramètres du modèle que l'on cherche à déterminer. Cette densité n'est néanmoins pas forcément sous une forme analytique, ni explicite. Celle-ci peut être fournie sous la forme d'une boîte noire que l'on évalue en certains points. Dans ce cas, la détermination des paramètres du modèle ne peut pas être réalisée par des méthodes dites classiques.

De plus, si l'on souhaite connaître, par exemple, l'espérance ou la variance de ces paramètres, il nous faudra résoudre des intégrales, qui, parfois, ne peuvent être résolues par des méthodes dites déterministes [10] ; cela est surtout le cas lorsque l'on traite des modèles avec un grand nombre de paramètres, ce qui donne lieu à des intégrales multiples. Une méthode pour résoudre ces intégrales s'appelle les méthodes MCMC (Markov Chain Monte Carlo).

Avant d'introduire ces méthodes, nous citons et illustrons, sur base de [2], deux méthodes classiques.

3.2 Méthode de l'inverse

La première méthode classique est assez limitée dans la pratique ; en effet, elle nécessite une expression analytique pour la densité de probabilité cible, et également que cette dernière soit inversible. Cette méthode fonctionne comme suit :

Soient

- F une fonction de répartition et U une variable aléatoire de loi uniforme sur $[0, 1]$.

– G une fonction définie par :

$$\begin{aligned} G : [0, 1] &\longrightarrow [0, 1] \\ u &\longmapsto G(u) = \inf\{y \mid F(y) \geq u\} \end{aligned}$$

Alors nous avons que G est une variable aléatoire dont la fonction de répartition vaut F .

Illustrons cette méthode par un exemple simple : considérons la loi exponentielle de fonction de densité suivante :

$$p(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

où $\lambda > 0$. On cherche à déterminer une variable aléatoire X dont la fonction de densité vaut $p(x, \lambda)$. On calcule la fonction de répartition de $p(x, \lambda)$:

$$\begin{aligned} F(x, \lambda) &= \int_{-\infty}^x p(x, \lambda) dx \\ &= \int_0^x p(x, \lambda) dx && \text{si } x \geq 0 \\ &= \int_0^x \lambda e^{-\lambda x} dx \\ &= 1 - e^{-\lambda x} \end{aligned}$$

On obtient :

$$F(x, \lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

En utilisant un générateur de nombres pseudo-aléatoires (fonction **rand**) suivant une loi uniforme sur $[0, 1]$, on peut créer une suite de nombres suivant la loi p en évaluant $F(\mathbf{rand}(), \lambda)$.

Les limites de cette méthode sont conséquentes : la fonction de densité doit être sous forme analytique et facilement intégrale, et sa fonction de répartition doit être inversible.

À la section suivante, nous introduisons une autre méthode classique, qui n'a pas besoin d'une forme explicite de la fonction cible, mais qui est limitée pour des densités de plusieurs variables.

3.3 Méthode du rejet

Considérons une densité de probabilité cible f , une densité quelconque g telle qu'il doit être facile d'engendrer une variable aléatoire X suivant g , un intervalle $[a, b]$ et une constante $M > 0$. Alors, si les deux points suivants sont satisfaits, la méthode du rejet permet de générer une suite de nombres suivant la loi de probabilité p sur $[a, b]$. Pour cela, il est nécessaire que :

1. $\forall x \notin [a, b] : p(x) = 0$ et $g(x) = 0$,
2. $\forall x \in [a, b] : p(x) \leq M g(x)$.

Afin de générer une suite de nombres suivant la densité cible, la méthode du rejet consiste à :

1. Générer x suivant la fonction de densité g , et u selon la distribution uniforme sur l'intervalle $[0, M]$.
2. Conserver x si $M u \leq \frac{f(x)}{g(x)}$, sinon recommencer à l'étape précédente.

Nous pouvons déjà noter que si M est trop grand (donc mal choisi), l'algorithme rencontrera beaucoup de rejets et sera donc moins performant.

Illustrons cette méthode afin d'engendrer une variable aléatoire suivant une loi f de probabilité gamma dans l'intervalle $[0, 10]$:

$$f(x) = x^{k-1} \frac{\beta^k e^{-\beta x}}{\Gamma(k)}, \quad k = 9, \beta = 2.$$

Nous allons utiliser une fonction de densité g suivant une loi normale de moyenne 4 et d'écart-type 2. Nous pouvons observer, à la FIGURE 3.1, que l'on a bien l'inégalité $f < M g$ pour $M = 1.5$. L'algorithme utilisé est présenté au LISTING 1. À la FIGURE 3.1, nous voyons que l'histogramme des nombres

obtenus semble bien être celui de la fonction de densité cible.

Comme précisé plus haut, cette méthode n'est pas la plus efficace pour des problèmes à un nombre de dimensions élevé. De plus, si la fonction g est mal choisie, un nombre important de rejets peut être rencontré, ce qui réduira les performances de l'algorithme. Notons toutefois qu'il est toujours possible de considérer une fonction de densité uniforme sur $[a, b]$ pour g .

Nous allons à présent présenter une méthode plus générale, qui peut donc être appliquée à n'importe quelle fonction de densité, avec moins de contraintes.

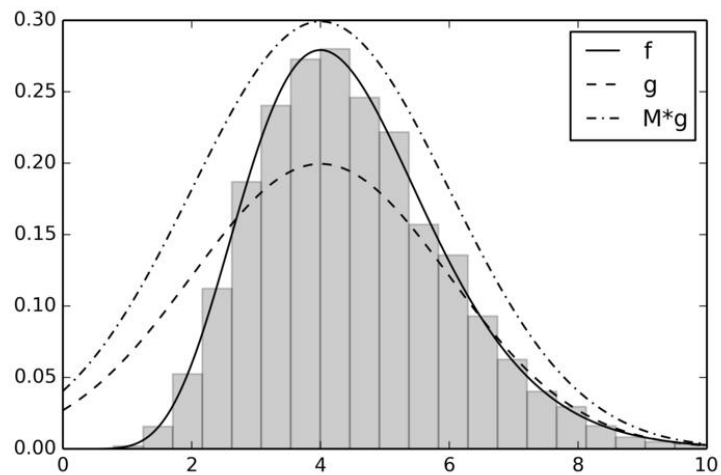


FIGURE 3.1 – Histogramme des nombres générés suivant la loi $f \sim \Gamma(9, 2)$ à l'aide de la méthode du rejet. Les fonctions de densité f , g et Mg sont également représentées.

```

# f: fonction de densite de la loi gamma(9,2)
# g: fonction de densite de la loi normale(4,2)
# grvs: generateur de nombre pseudo-aleatoires suivant g
xmin, xmax, M = 0, 10, 1.5
def genprand(a, b, M, fpdf, g, grvs):
    while True:
        x = grvs()
        if not(xmin < x < xmax):
            continue
        u = rand()*M
        if u*M*g(x) < fpdf(x):
            yield x
p = genprand(-2, 2, M, f, g, grvs)
samples = []
for i in xrange(10000):
    samples.append(p.next()) # genere 10000 nbres suivant g

```

Listing 1: Méthode du rejet appliquée à la loi $\Gamma(9, 2)$.

3.4 Méthode de Monte-Carlo par chaînes de Markov

Cette partie a pour objectif de présenter l'algorithme de Metropolis-Hastings. Comme celui-ci est basé sur le concept de marcheurs aléatoires, nous commencerons par définir ces derniers.

Dans un premier temps, nous définissons les chaînes de Markov sur base d'un exemple. Ensuite, nous listons les propriétés qui conduisent à un théorème de convergence des chaînes vers une distribution d'équilibre.

Enfin, nous détaillons l'algorithme de MH, montrons qu'une chaîne de Markov générée par ce dernier atteint bien sa distribution d'équilibre, et l'illustrerons sur deux densités cibles à une et à deux dimensions.

3.4.1 Chaînes de Markov

3.4.1.1 Définitions

Nous introduisons donc, dans cette section, les définitions principales liées aux chaînes de Markov, sur base de l'exemple suivant.

On se place dans le cadre d'une ville constituée de quatre rues formant les côtés un carré $ABCD$. On nomme les angles de la rue $(A, B, C, D) = (1, 2, 3, 4)$ et on suppose qu'une personne se déplace d'un angle à l'autre. Au départ le marcheur se trouve en 1, il lance une pièce équilibrée pour savoir quelle rue emprunter : il peut aller soit en 2, soit en 4. Après s'être déplacé la première fois, il lance à nouveau la pièce : si c'est pile il se rend à l'angle suivant dans le sens horloger, si c'est face il se rend à l'angle suivant dans le sens « anti-horloger ». Il continue de marcher en répétant le même processus.

On considère les VA $(X_n)_{n \in \mathbb{N}}$ tq $X_i \equiv$ “Position du marcheur au temps t_i ” ; on connaît la loi de probabilité de X_0 et X_1 :

$$\begin{aligned} P[X_0 = 1] &= 0 \\ P[X_1 = 2] &= 0.5 \\ P[X_1 = 4] &= 0.5 \\ P[X_0 = i] &= P[X_1 = j] = 0 \quad \forall i \neq 1, \forall j \notin \{1, 4\} \end{aligned}$$

Afin de déterminer la loi de probabilité pour $n > 1$, on doit passer par les probabilités conditionnelles : on suppose au temps t_n que le marcheur se trouve à la position 2. La probabilité qu'il se situe à une position au temps suivant s'exprime par :

$$\begin{aligned} P[X_{n+1} = 1 | X_n = 2] &= 0.5 \\ P[X_{n+1} = 3 | X_n = 2] &= 0.5 \\ P[X_{n+1} = i | X_n = 2] &= 0 \quad \forall i \notin \{1, 3\} \end{aligned}$$

On peut remarquer que seul l'état courant est nécessaire pour connaître l'état

suivant, $\forall \alpha_0, \dots, \alpha_n$:

$$P[X_{n+1} = 1 | X_n = 2, X_{n-1} = \alpha_{n-1}, \dots, X_2 = \alpha_2, X_1 = \alpha_1, X_0 = \alpha_0] = 0.5$$

Ces variables aléatoires forment ce que l'on appelle une chaîne de Markov. Nous les définissons comme suit :

Définition 3.1 Chaîne de Markov

Soit (X_n) une suite de variables aléatoires telle que, $\forall n$, $X_n : \Omega \longrightarrow \mathbb{R}$ où $X(\Omega) = \{s_1, \dots, s_k\} = \{s_i\}_{i \in J}$. Cette suite définit une chaîne de Markov ssi

$$P[X_{n+1} = s_j | X_n = s_i, X_{n-1} = s_{i_{n-1}}, \dots, X_0 = s_{i_0}] = P[X_{n+1} = s_j | X_n = s_i]$$

On peut remarquer, dans l'exemple, que $P[X_{n+1} = i | X_n = 2] = \alpha$, où α est indépendant de n . Quand une chaîne de Markov ne dépend pas de l'état n , elle est dite homogène. Ces chaînes homogènes sont définies formellement comme suit :

Définition 3.2 Chaîne de Markov homogène

Une chaîne de Markov est dite **homogène** ssi $\forall n$: $P[X_{n+1} = s_j | X_n = s_i] = \alpha_{ij}$. On note cette probabilité P_{ij} que l'on appelle **probabilité de transition** de la chaîne de Markov.

Une chaîne de Markov homogène est déterminé par $(P_{ij})_{i,j \in J}$, appelée **matrice de transition**. On représente la probabilité initiale de X_0 sous la forme d'un vecteur noté $\mu^{(0)}$:

$$\mu^{(0)} = (P[X_0 = s_1], \dots, P[X_0 = s_k])$$

Dans cet exemple, on a :

$$\mu^{(0)} = (P[X_0 = 1], \dots, P[X_0 = 4]) = (1, 0, 0, 0)$$

et la probabilité de transition vaut donc :

$$(P)_{ij} = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

Si l'on note de manière générale $\mu^{(n)}$ la probabilité au temps n comme suit :

$$\mu^{(n)} = (P[X_n = s_1], \dots, P[X_n = s_k]).$$

Nous pouvons montrer que l'égalité (3.1) suivante est vérifiée :

$$\mu^{(n)} = \mu^{(0)} P \quad (3.1)$$

En effet, notons chaque élément j du vecteur $\mu^{(n)}$ par : $\mu_j^{(n)} = P[X_n = s_j]$. En posant $B_i = \{\omega \in \Omega \mid X_{n-1}(\omega) = s_i\}$, comme $X_{n-1}(\Omega) = \{s_1, \dots, s_k\}$, on a : $\cup_{i \in J} B_i = \Omega$; donc $(B_i)_{i \in J}$ forme une partition de Ω . On peut appliquer la formule des probabilités totales :

$$\begin{aligned} P[X_n = s_j] &= \sum_{i=1}^k P[X_n = s_j \mid X_{n-1} = s_i] P[X_{n-1} = s_i] \\ &= \sum_{i=1}^k P_{ij} \mu_i^{n-1} \end{aligned} \quad (3.2)$$

Sous forme matricielle, l'équation (3.2) s'écrit $\mu^{(n)} = \mu^{(n-1)} P$. Par récurrence sur n , on a :

$$\mu^{(n-1)} = \mu^{(n-2)} P = \dots = \mu^{(0)} P.$$

Nous avons défini ce que sont les chaînes de Markov, et démontré une propriété importante de la matrice de transition des chaînes homogènes.

Cette dernière sera utilisée à la section suivante, où nous définissons le concept de distribution d'équilibre d'une chaîne de Markov.

3.4.1.2 Distribution d'équilibre ou stationnaire

Nous introduisons dans cette partie une notion importante des chaînes de Markov qui est la **distribution d'équilibre**. Nous illustrons le concept par un exemple avant d'en donner la définition formelle.

Soit un problème qui consiste à prédire le temps du jour n en fonction du jour $n - 1$:

- S'il pleut, il y aura du soleil le lendemain avec une probabilité de 0.5.
- S'il y a du soleil, il pleuvra le lendemain avec une probabilité de 0.1.

On remarque que le temps qu'il fera ne dépend que du temps qu'il fait la veille. On définit les variables aléatoires X_n comme "le temps qu'il fait le n -ième jour". $X_n(\Omega) = \{s_1, s_2\}$ où s_1 correspond à un temps de pluie et s_2 au soleil. La probabilité de transition vaut :

$$\begin{aligned} (P)_{ij} &= \begin{pmatrix} P[X_n = s_1 | X_{n-1} = s_1] & P[X_n = s_2 | X_{n-1} = s_1] \\ P[X_n = s_1 | X_{n-1} = s_2] & P[X_n = s_2 | X_{n-1} = s_2] \end{pmatrix} \\ &= \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{pmatrix} \end{aligned}$$

On pose $\mu^{(0)} = (1, 0)$. Par l'équation (3.1), on a :

$$\begin{aligned} u^{(1)} &= u^{(0)} P = (1, 0) P = (0.5, 0.5) \\ u^{(2)} &= u^{(0)} P^2 = (0.3, 0.7) \\ &\vdots \\ u^{(n)} &= u^{(n-1)} P^n \end{aligned}$$

Nous avons démontré, à l'annexe A, que le vecteur $u^{(n)}$ converge vers $(\frac{5}{6}, \frac{1}{6})$. On peut également remarquer ([10]) qu'avec une condition initiale $\mu^{(0)}$ différente, la chaîne $\mu^{(n)}$ aurait convergé vers la même distribution, qui est alors appelée *distribution d'équilibre*.

On définit formellement la distribution d'équilibre d'une chaîne de Markov comme suit :

Définition 3.3 Distribution d'équilibre

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov dont les variables aléatoires sont à valeurs dans $\{s_1, \dots, s_k\}$, et soit P une matrice de transition. On a alors qu'un vecteur ligne $\pi = (\pi_1, \dots, \pi_k)$ tq $\pi_i \geq 0 \ \forall i = 1, \dots, k$ et $\prod_{i=1}^k \pi_i = 1$ est appelé distribution d'équilibre ssi $\pi = \pi P$.

Remarque : La définition est cohérente : si $u^{(n)} = \pi$, alors $u^{(n+1)} = u^{(n)}P = \pi P = \pi$. D'où, par récurrence sur n :

$$u^{(k)} = \pi \quad \forall k \geq n.$$

Maintenant que nous savons ce qu'est une distribution d'équilibre pour une chaîne de Markov, nous pouvons introduire le théorème de convergence. Celui-ci impose plusieurs propriétés sur les chaînes, que nous détaillons également à la section suivante.

3.4.1.3 Propriétés

Il est possible d'imposer certaines propriétés sur une chaîne de Markov pour que celle-ci converge vers une distribution stationnaire. Ces propriétés sont : l'irréductibilité, la récurrence, la récurrence positive et l'a-périodicité d'une chaîne. Sur base du troisième chapitre de W.R. Gilks *et al* [5], nous allons définir ces propriétés et citer le théorème exprimant cette convergence.

Introduisons d'abord une nouvelle notation pour la probabilité de transition de $X_t = j$ à $X_0 = i$ comme suit :

$$P_{ij}(t) = P[X_t = j | X_0 = i].$$

Propriété 3.1 Chaîne irréductible

Une chaîne X est dite **irréductible** si elle peut atteindre un état de probabilité non nulle en un nombre fini d'itérations, i.e. ssi

$$\forall i, j, \exists t > 0 : P_{ij}(t) > 0$$

Propriété 3.2 Chaîne récurrente

Une chaîne qui, partant d'un état i , atteindre sûrement à nouveau cet état en un nombre d'itération finie, est dite irréductible et **récurrente**; i.e. X est **récurrente** ssi

- X est irréductible,
- $\forall i : P_{ij}(\tau_{ii} < \infty)$,

où $\tau_{ii} = \min \{t > 0 : (X_t = i | X_0 = i)\}$. Une chaîne non récurrente est dite **transiente**.

Propriété 3.3 Chaîne récurrente positive

Une chaîne récurrente X est dite **récurrente positive** s'il existe une distribution d'équilibre pour X . Avec notre notation $P_{ij}(t)$, cela s'exprime comme suit : X est récurrente positive ssi X est récurrente et

$$\exists \pi(\cdot), j, t \geq 0 : \sum_i \pi(i) P_{ij}(t) = \pi(j). \quad (3.3)$$

Propriété 3.4 Chaîne apériodique

Une chaîne irréductible X partant de l'état i et revenant plusieurs fois à cet état de manière non régulier est dite apériodique; i.e. X est apériodique ssi

- X est irréductible,
- $1 = \gcd\{t > 0 : P_{ii}(t) > 0\}$,

où \gcd représente le « plus grand commun diviseur ».

L'intégrale $I = \int h(x)p(x)dx$, où $p(x)$ est une densité de probabilité, peut être approchée par : $\overline{f}_n = \frac{1}{n} \sum_{t=1}^n f(X_t)$, où la variable aléatoire X_t est engendrée par la loi de densité $p(x)$. Le théorème suivant exprime la convergence de toutes chaînes de Markov **ergotiques** (i.e. apériodique et récurrente positive) vers sa distribution stationnaire.

Théorème 3.1 Soit X une chaîne de Markov ergotique, alors sa distribution d'équilibre est l'unique distribution de probabilité satisfaisant (3.3). On a alors, de plus :

- $\forall i, j : P_{ij}(t) \rightarrow \pi(j)$ quand $t \rightarrow \infty$,
- Si $\sum_i |f(x_i)|\pi(i) < \infty$, alors $P[\overline{f}_n \rightarrow \sum_i f(i)\pi(i)] = 1$ pour $n \rightarrow \infty$.

Maintenant que nous avons détaillé les principaux concepts liés aux chaînes de Markov, nous pouvons introduire l'algorithme de Metropolis-Hastings. Nous vérifierons que les chaînes produites par ce dernier convergent bien vers une distribution stationnaire.

3.4.2 Algorithme de Metropolis-Hastings

Nous allons détailler les étapes principales de l'algorithme de Metropolis-Hastings (MH), montrer que les chaînes de Markov générées par lui convergent bien vers leur distribution stationnaire. Ensuite, nous l'appliquerons sur deux exemples de densité de probabilité à une et à deux dimensions.

3.4.2.1 Étape principale de l'algorithme

On désire créer une réalisation d'une chaîne de Markov dont la fonction d'équilibre vaut $\pi(x)$. On suppose qu'à l'itération n , le programme a généré $n - 1$ valeurs de X_1, \dots, X_n . À l'itération $n + 1$, l'algorithme a uniquement besoin de la valeur $X_n = x$ pour déterminer $X_{n+1} = x'$. La valeur de x' est déterminée comme suit :

- On génère x' à partir d'une fonction de densité $g(x'|x)$ (il doit être facile d'engendrer des échantillons suivant cette fonction).
- On accepte x' avec une probabilité $\alpha(x'|x)$ définie par :

$$\alpha(x'|x) = \min \left\{ 1, \frac{\pi(x') g(x|x')}{\pi(x) g(x'|x)} \right\}.$$

- Si x' est refusé, alors $X_{n+1} = x$.

Remarquons qu'il n'est pas nécessaire d'avoir une formulation de la distribution $\pi(x)$ qu'à un facteur multiplicatif près. Cette propriété de l'algorithme avait été citée à la section 2.2 lors de l'introduction à la méthode de l'inférence bayésienne.

Montrons à présent la convergence des chaînes générées par l'algorithme vers leur distribution stationnaire.

3.4.2.2 Convergence vers une distribution d'équilibre

Nous pouvons vérifier¹ que l'algorithme de Metropolis-Hastings génère bien une chaîne de Markov dont la distribution d'équilibre est π . Pour ce faire, considérons cette fois un espace d'états continu $\Omega \in \mathbb{R}$. La probabilité $p(x'|x)$ (appelée *noyau* de la chaîne de Markov) est définie alors par

$$P[x' \leq X_{n+1} < x' + dx | X_n = x].$$

On alors alors :

$$\int_{\Omega} p(x'|x) dx' = 1. \quad (3.4)$$

En effet :

$$\int_{\Omega} p(x'|x) dx' = \frac{1}{p(x)} \int_{\Omega} p(x', x) dx' \quad (3.5)$$

$$= \frac{1}{p(x)} p(x) \quad (3.6)$$

$$= 1,$$

où (3.5) est obtenue par définition de probabilité conditionnelle, et (3.6) par la définition de fonction marginale.

Afin que l'algorithme à l'étape $n + 1$ génère x' , ($x' \neq x$), il faut que celui-ci soit généré par la $g(x', x)$ et qu'il soit accepté. On a donc :

$$p(x'|x) = g(x'|x) \alpha(x'|x).$$

Montrer que π définit une distribution d'équilibre dans le cas continu, se généralise à partir de l'équation (3.3) par :

$$\int_{\Omega} \pi(x) p(x'|x) dx = \pi(x').$$

1. Sur base de Koot [10] section 2.2.3 et W.R. Gilks *et al* [5] section 1

Ce qui est bien le cas ici, en effet :

$$\begin{aligned}
 p(x'|x)\pi(x) &= \alpha(x'|x)g(x'|x)\pi(x) \\
 &= \min \left\{ 1, \frac{\pi(x')g(x|x')}{\pi(x)g(x'|x)} \right\} g(x'|x)\pi(x) \\
 &= \min \{ \pi(x)g(x'|x), \pi(x')g(x|x') \} \\
 &= \min \left\{ 1, \frac{\pi(x)g(x'|x)}{\pi(x')g(x|x')} \right\} g(x|x')\pi(x') \\
 &= \alpha(x|x')g(x|x')\phi(x') \\
 &= p(x|x')\phi(x')
 \end{aligned}$$

En intégrant par rapport à x , on obtient par (3.4) :

$$\int_{\Omega} p(x'|x)\pi(x) = \pi(x').$$

Notons ici que l'on a simplement montré la convergence vers une distribution stationnaire, mais que l'on n'a pas démontré que les chaînes générées par l'algorithme respectent bien les propriétés nécessaires à l'application du théorème 3.1.

Les deux sections suivantes serviront à illustrer l'algorithme dans des exemples concrets.

3.4.2.3 Sampling selon une loi à une dimension

Appliquons l'algorithme de Metropolis-Hastings à un exemple concret à une dimension. Soit la probabilité cible $p \sim N(0, 1)$. Afin de générer une suite de markov dont la distribution d'équilibre convergera vers f , il nous faut définir une densité de proposition $g(x', x)$. Nous choisissons la loi uniforme centrée autour de $\delta = 0.5$, i.e. :

$$g(x'|x) = \begin{cases} \frac{1}{2\delta} & \text{si } x - \delta < x' < x + \delta, \\ 0 & \text{sinon.} \end{cases}$$

L'algorithme est présenté au LISTING 2, et le graphe des résultats à la FIGURE 3.2. On constate que l'histogramme des samples générés semble bien être celui de la loi de densité cible. Nous pouvons cependant améliorer l'algorithme :

- Par symétrie de la densité de proposition, i.e. $g(x', x) = g(x, x')$, la probabilité $\alpha(x'|x)$ pour accepter x' devient :

$$\alpha(x'|x) = \min \left\{ 1, \frac{\pi(x')}{\pi(x)} \right\}.$$

L'algorithme est alors l'original de Metropolis.

- Au lieu d'utiliser les fonctions de densité telles quelles, on peut prendre leur logarithme. Cela permettrait de ne pas avoir d'erreur de division par zéro lors du test pour accepter x' ; en effet, on aurait alors :

$$\alpha(x'|x) = \min \{ 0, \pi(x') - \pi(x) + g(x|x') - g(x'|x) \}$$

Notons déjà que les deux librairies que nous utiliserons lors de nos tests exigent des fonctions de probabilité sous leur forme logarithmique.

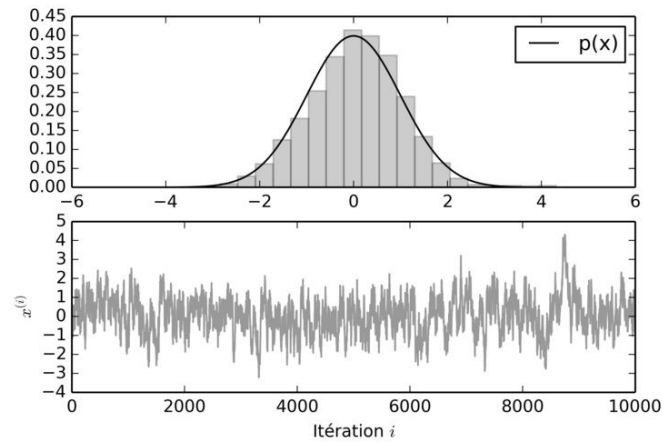


FIGURE 3.2 – Simulation de la loi $p(x)$ par l’algorithme de Metropolis-Hastings. Le graphe supérieur représente la fonction de densité cible ainsi que l’histogramme des échantillons générés par l’algorithme. Le graphe inférieur représente les échantillons acceptés au cours des itérations.

```

# p(x) fct de densite cible
# g(x',x) densite de proposition
# qrvs(x) genere nbres selon la loi g
n, mu, sigma, delta = 10000, 0, 1, 0.5
x0 = qrvs(0)
samples = [x0]
for i in xrange(n):
    xprev = samples[-1]
    xnew = qrvs(xprev)
    pprev, pnew = p(xprev), p(xnew)
    qprev, qnew = q(xprev, xnew), q(xnew, xprev)
    a = min(1, (pnew/pprev) * (qprev/qnew))
    u = np.random.uniform()
    if u < a:
        samples.append(xnew)
    else:
        samples.append(xprev)

```

Listing 2: Algorithme de Metropolis-Hastings appliqué pour simuler une fonction de densité suivant une loi normale $\mathcal{N}(0, 1)$ avec une densité de proposition $g(x', x)$ uniforme centrée autour de $\delta = 0.5$.

3.4.2.4 Sampling selon une loi à deux dimensions

Nous allons maintenant, sur base de « Homework 5 » de [13], simuler une loi de probabilité à deux dimensions $p(x, y)$ définie comme suit :

$$p(x, y) = \exp \left(-100 \left(\sqrt{y^2 + x^2} - 1 \right)^2 + (x - 1)^2 - y - 5 \right)$$

Nous utiliserons une densité de proposition $g(x'|x)$ suivant une loi normale centrée en x et d'écart-type 0.5, i.e. :

$$g(x'|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x'-x}{\sigma}\right)^2}, \quad \sigma = 0.5.$$

L'algorithme de Metropolis-Hastings se généralise directement pour des fonctions à plusieurs dimensions, mais ce n'est pas ce qui est appliqué dans la pratique. En effet, cela reviendrait à générer un vecteur $\mathbf{x}' = (x'_1, \dots, x'_n)$, puis de l'accepter ou de le rejeter suivant $\alpha(\mathbf{x}'|\mathbf{x})$. Cela peut donner lieu à beaucoup trop de rejets, ce qui fait que l'algorithme convergera en un temps plus long. La méthode utilisée généralement est d'accepter ou de rejeter chaque élément de \mathbf{x}' consécutivement. À chaque étape de l'algorithme, la valeur de \mathbf{x}' en fonction de \mathbf{x} est déterminée comme suit :

1. Générer x'_1 suivant $g(x'_1|x_1)$,
2. Accepter x'_1 avec une probabilité $\alpha(x'_1, x_1) = \min \left\{ 1, \frac{p(x'_1, x_2)}{p(x_1, x_2)} \frac{g(x_1|x'_1)}{g(x'_1|x_1)} \right\}$,
3. Poser $\tilde{x}_1 = x'_1$ si x'_1 est accepté, sinon $\tilde{x}_1 = x_1$,
4. Générer x'_2 suivant $g(x'_2|x_2)$,
5. Accepter x'_2 avec une probabilité $\alpha(x'_2, x_2) = \min \left\{ 1, \frac{p(\tilde{x}_1, x'_2)}{p(\tilde{x}_1, x_2)} \frac{g(x_2|x'_2)}{g(x'_2|x_2)} \right\}$,
6. Poser $\tilde{x}_2 = x'_2$ si x'_2 est accepté, sinon $\tilde{x}_2 = x_2$,
7. $\mathbf{x}' = (\tilde{x}_1, \tilde{x}_2)$.

Le code utilisé figure au LISTING 3, celui-ci s'adapte à des problèmes à un nombre de dimensions quelconque. L'algorithme de Metropolis-Hastings génère deux chaînes de Markov, pour x et y . La chaîne pour x suit la loi de densité marginale $p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy$, de même pour y . On peut le vérifier à la FIGURE 3.3. On constate bien que, pour chaque chaîne, l'histogramme

semble bien être celui d'une densité correspond à la densité marginale en question.

Maintenant que nous avons un algorithme permettant d'engendrer une variable aléatoire suivant une densité de probabilité quelconque, nous pouvons appliquer la méthode de l'inférence bayésienne dans des cas concrets. On illustrera cette méthode au chapitre suivant via deux exemples. De plus, nous parlerons des outils utilisés pour réaliser ces tests.

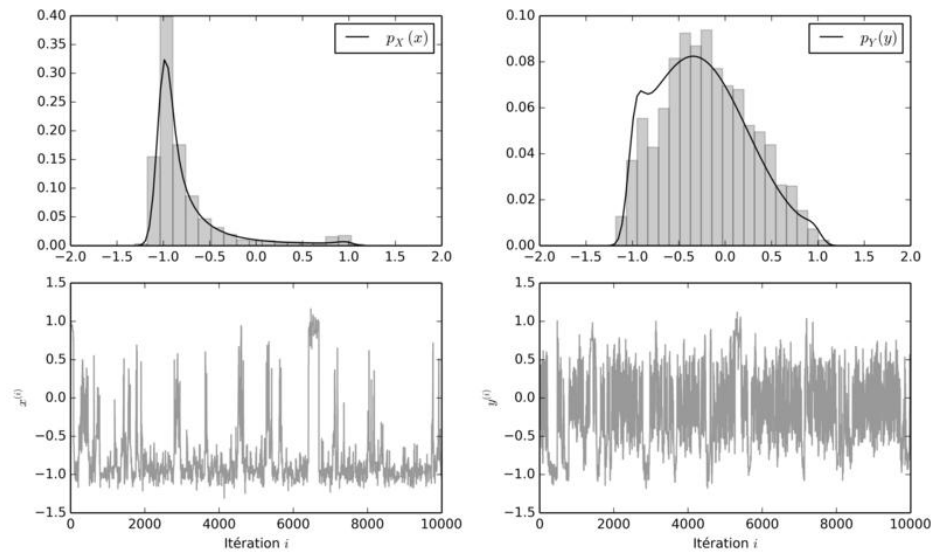


FIGURE 3.3 – Résultats de l'algorithme de Metropolis-Hastings pour une fonction de densité $p(x, y)$. Les graphes de gauche concernent la variable x , et ceux de droite la variable y . Le graphe supérieur gauche représente la densité marginale $p_X(x)$ et l'histogramme des chaînes de Markov générés pour x . Le graphe inférieur gauche trace les samples $x^{(i)}$ au cours des itérations. (Graphes de droites similaires selon y).

```
# p(x,y) densite cible
# q(x',x) densite de proposition g(x'/x)
# qrvs(x) genere des nombres suivant g(x'/x)
n, dim, sigma = 10000, 2, 0.5
samples = ndarray((n, dim))
samples[0] = random.normal(0, sigma, size=2)
for i in xrange(n-1):
    samples[i+1] = samples[i]
    for j in xrange(dim):
        samples[i+1,j] = qrvs(samples[i+1,j])
        pnew, pprev = p(*samples[i+1]), p(*samples[i])
        qnew = q(samples[i+1,j], samples[i,j])
        qold = q(samples[i,j], samples[i+1,j])
        u = np.random.uniform()
        a = min(1, (pnew/pprev)*(qold/qnew))
        if a <= u: # rejet
            samples[i+1,j] = samples[i,j]
```

Listing 3: Algorithme de Metropolis-Hastings appliquée à une fonction de densité à deux dimensions. Généralisable en un nombre quelconque de dimensions.

Chapitre 4

Applications introductives sur l'inférence bayésienne

Dans ce chapitre, nous appliquerons la méthode de l'inférence bayésienne à deux problèmes relativement simple : le premier est un problème de régression linéaire, où l'on obtiendra un ensemble d'échantillons probables représentant la pente de la droite recherchée. Le deuxième consiste à déterminer l'origine d'un événement sismique, sachant que des stations aux alentours ont noté le moment où elles ont perçu l'onde sismique.

4.1 Outils

Les résultats obtenus dans ce chapitre ont été calculés numériquement de trois manières : en utilisant un algorithme MCMC personnel, avec la librairie python PYMC¹, et avec la librairie python EMCEE. Notons que les résultats publiés dans ce mémoire sont le résultat de calculs réalisés avec EMCEE. En effet, nous l'avons préféré aux autres car il gère le multi-threading, est codé en python pur (donc il s'installe facilement) ; de plus, EMCEE est la librairie sur laquelle se base COSMOHAMMER², un framework destiné essentiellement à l'estimation de paramètres cosmologiques via MCMC. Un article a également été publié [6] sur arXiv expliquant, notamment, le choix des fonctions de propositions, requises par l'algorithme de Metropolis-Hastings, pris par les auteurs.

4.2 Régression linéaire

Nous allons ici réaliser une régression linéaire en utilisant l'inversion bayésienne et les méthode MCMC. La formalisation du problème et l'exemple de cette partie se base sur M. West [16] et P. Protopapas [13].

Soient $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ où x est fixé et y sont les données observées. On souhaite estimer a tel que

$$y_i = ax_i + b + \epsilon_i,$$

où $b = 0$ et ϵ_i sont des erreurs de mesures suivant une gaussienne : $\epsilon_i \sim N(0, \sigma)$.

Le principe de l'inférence bayésienne consiste à exprimer la fonction a posteriori $p(\theta|d)$, où, ici, $d = y$ et $\theta = a$. Comme les erreurs suivent une loi gaussienne, on a comme fonction de vraisemblance (via (2.7)) :

$$L(y_i|a) = \prod_{i=1}^m \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[-\sum_{i=1}^m \frac{1}{2} \left(\frac{d_i - M_i(a)}{\sigma} \right)^2 \right],$$

1. <https://pymc-devs.github.io/pymc/>

2. <http://www.astro.ethz.ch/refregier/research/Software/cosmohammer>

où $M_i(a) = x_i a$. Par l'équation (2.1), la probabilité a posteriori recherchée est : $p(a|y) = \frac{L(y|a)\pi(a)}{\int_{\mathcal{A}} L(y|a)\pi(a)da}$. Comme l'algorithme de metropolis-hastings ne nécessite qu'une fonction à un facteur multiplicatif près, nous ne considérons que le numérateur comme fonction de densité cible.

4.2.1 Données du problème

Nous allons nous-même créer les données du système. Nous posons d'abord : $n = 9$, $\sigma = 1.5$, $a = 1.2$, et $b = 0$. On génère ensuite x selon une loi uniforme à valeur dans l'intervalle $[-3, 7]$. Enfin, on génère les données y telles que $y_i = ax_i + \sigma\epsilon_i$, $\forall i$, où ϵ_i sont choisis aléatoirement selon une loi normale centrée en 0 et d'écart-type σ .

4.2.2 Résolution

La likelihood étant la fonction de vraisemblance représentée plus haut, il nous faut également définir une prior. Celle-ci permet de prendre en considération les informations dont on dispose sur a . Nous choisirons la densité d'une loi normale centrée en 0 et d'écart-type 4.

Les graphes des fonctions de densités de la likelihood et de la prior sont représentés à la FIGURE 4.1. L'histogramme du paramètre a obtenu se trouve à la FIGURE 4.2. On constate que les samples a_i semblent suivre une loi normale de moyenne 1.2976 et d'écart-type 0.1223. Nous le vérifions grâce à un test d'ajustement de Kolmogorov-Smirnov. On obtient une *p-valeur* de 0.074, signifiant, pour un taux d'erreur de première espèce $\alpha = 0.5$, que l'on rejette pas l'hypothèse nulle selon laquelle a_i suivent une loi normale telle que spécifiée.

Nous pouvons également constater que les samples a_i obtenus sont bien ceux qui maximisent la likelihood (sur le graphe de droite de la FIGURE 4.1).

Le graphe des résultats est illustré à la FIGURE . La droite pleine désigne le a initial (sans tenir compte des erreurs d'observations $y * i$), celle en pointillé

correspond à la droite obtenue en prenant la moyenne des samples obtenus comme a . Enfin, la partie grisée correspond à l'ensemble des droites probables (i.e. obtenues en fonction de chaque samples a_i des chaînes de markov générées par l'algorithme de Metropolis-Hastings).

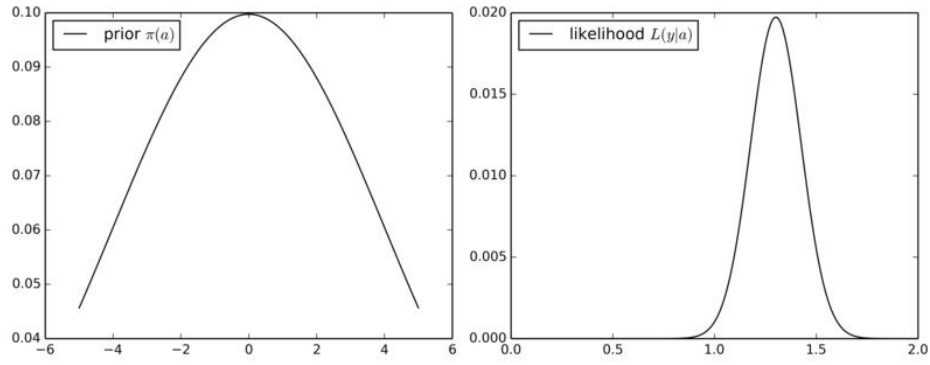


FIGURE 4.1 – Fonction de densités de la likelihood et de la prior pour la résolution d'un problème de régression linéaire du type $y_i = ax_i$ par inférence bayésienne.

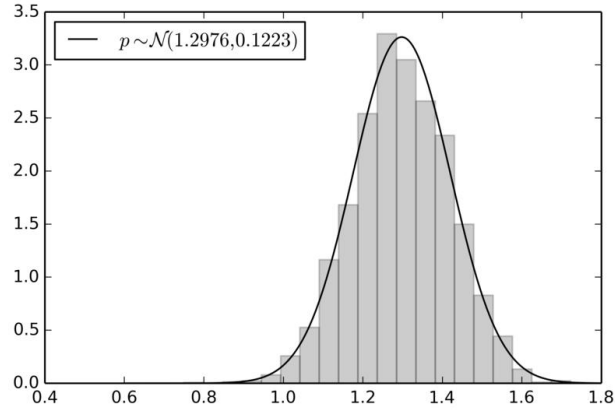


FIGURE 4.2 – Histogramme des samples a_i obtenus lors de la résolution d'un problème de régression linéaire du type $y_i = ax_i$. Représentation également de la loi de densité normale que les samples a_i semblent suivre selon le test de Kolmogorov-Smirnov pour un taux d'erreur de première espèce $\alpha = 0.5$.

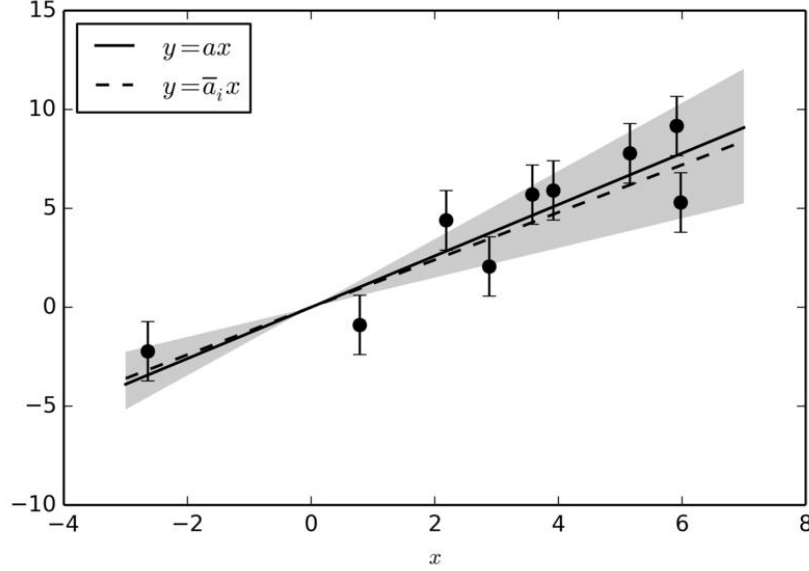


FIGURE 4.3 – Solution au problème de régression linéaire $y = ax$ obtenu par inférence bayésienne. La droite $y = ax$ représente la solution pour a initial en ne tenant pas compte des erreurs. La zone grise représente toutes les solutions probables obtenues par inférence bayésienne, où la droite moyenne est tracée selon $y = \bar{a}_i x$, où \bar{a}_i représente la moyenne des N samples générés par l'algorithme de Metropolis-Hastings, i.e. $\bar{a}_i = \frac{1}{N} \sum_{i=1}^N a_i$.

4.3 Épicentre d'un événement sismique

Nous allons dans cette section, sur base de C. Koot [10], estimer par inférence bayésienne l'épicentre d'un séisme. Nous disposons, comme données, de six stations (située en (x_i, y_i)) qui mesurent le temps d'arrivée t_i des ondes sismiques. Celles-ci sont répertoriées au tableau 4.1. On supposera que les temps d'arrivées suivent une erreur gaussienne d'écart-type $\sigma = 0.1$. On notera $v = 5 \text{ m s}^{-1}$ la vitesse de l'onde et t le moment où l'événement a eu lieu. On désire donc connaître le lieu (X, Y) de l'événement sismique, et le moment T où il a eu lieu, en inférant sur les données dont on dispose. Le problème d'inférence bayésienne s'exprime comme suit :

$$p(X, Y, T | \mathbf{x}, \mathbf{y}, \mathbf{t}) \propto p(\mathbf{x}, \mathbf{y}, \mathbf{t} | X, Y, T) p(\mathbf{x}, \mathbf{y}, \mathbf{t}),$$

	x_i	y_i	t_i
station 1	3	15	3.06
station 2	3	16	3.30
station 3	4	15	2.91
station 4	4	16	2.97
station 5	5	15	2.82
station 6	5	16	2.94

TABLE 4.1 – Coordonnées des différentes stations et moment à laquelle chacune a mesuré l'arrivée d'une onde sismique.

où $\mathbf{t} = t_1, \dots, t_n$, $\mathbf{x} = x_1, \dots, x_n$ et $\mathbf{y} = y_1, \dots, y_n$.

Nous procéderons en deux étapes : d'abord en considérant que le moment du séisme est connu et poserons $T = 0$, puis en le rajoutant comme paramètre inconnu du système suivant une gaussienne centrée en $\mu_T = 0.15$ et d'écart-type $\sigma_T = 0.2$. Nous n'imposerons pas de conditions sur les paramètres X et Y , i.e. leur prior suit une loi uniforme sur un intervalle suffisamment grand.

4.3.1 Moment de l'événement sismique connu

On suppose donc que $T = 0$. Comme les erreurs des données suivent une gaussienne, on obtient la likelihood suivante :

$$\mathcal{L}(X, Y) = \frac{n}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - M_i(X, Y)}{\sigma} \right)^2 \right], \quad n = 6, \sigma = 0.1.$$

Le modèle consiste à déterminer le moment \tilde{t}_i où la station i reçoit l'onde sismique, sachant le point de départ de l'événement sismique. On a $d_i = v\tau_i$, où d_i est la distance parcourue par l'onde entre (X, Y) et (x_i, y_i) et τ_i le temps qu'a pris l'onde pour parcourir la distance d_i . On peut écrire $\tau_i = \tilde{t}_i - T$, on a alors :

$$\tilde{t}_i = \frac{1}{v}d_i + T,$$

où T est ici nul. Comme la distance d_i représente la distance entre deux points (X, Y) et (x_i, y_i) , le modèle s'exprime finalement par :

$$M_i(X, Y) = \frac{1}{v} \sqrt{(X - x_i)^2 + (Y - y_i)^2}. \quad (4.1)$$

Les prior choisies étant des lois uniformes sur un intervalle suffisamment grand, celles-ci sont donc proportionnelles à 1. La fonction cible a posteriori correspond donc à la likelihood. Nous appliquons l'algorithme de Metropolis-Hastings pour inférer sur les paramètres X et Y . Les résultats sont présentés à la FIGURE 4.4. X et Y sont distribués autour des moyennes respectives 14.86 et 6.43. Nous constatons que les deux variables semblent être corrélées.

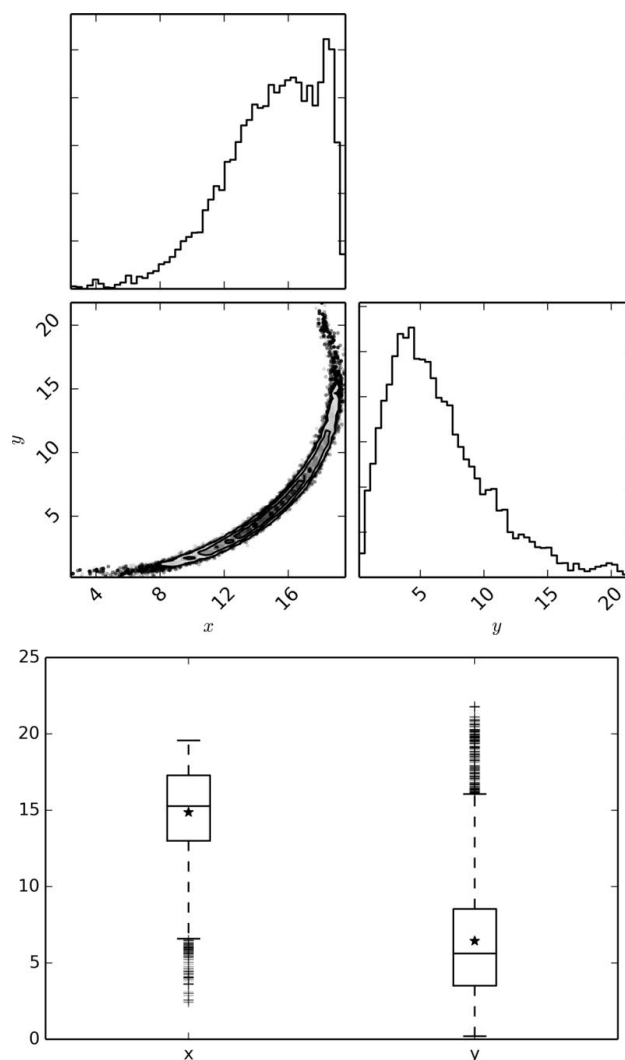


FIGURE 4.4 – Résultats de la détermination de l'endroit d'un événement sismique via inférence bayésienne, où l'on suppose que le moment du séisme est connu. Le graphe du haut montre les histogrammes des positions probables x et y du séisme, ainsi qu'un histogramme à deux dimensions groupant ces positions. Le graphes du bas est un boxplots des samples x et y obtenus.

4.3.2 Moment de l'événement sismique inconnu

Nous reprenons le même schéma qu'au point précédent, mais en considérant le moment initial de l'événement sismique comme paramètre supplémentaire au modèle. Comme indiqué précédemment, celui-ci suit une loi normale. On

a donc la prior suivante :

$$\pi(T) = \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left[-\frac{1}{2}\left(\frac{T - \mu_T}{\sigma_T}\right)^2\right], \quad \mu_T = 0.15, \sigma_T = 0.2.$$

Nous reprenons le même modèle qu'à l'équation (4.1) en considérant le paramètre T comme n'étant pas forcément nul. On obtient :

$$M_i(X, Y, T) = \frac{1}{v} \sqrt{(X - x_i)^2 + (Y - y_i)^2} + T.$$

La likelihood est identique mais prend compte de T dans le modèle, et la fonction a posteriori devient :

$$p(X, Y, T | \mathbf{x}, \mathbf{y}, \mathbf{t}) \propto \mathcal{L}(X, Y, T) \pi(T).$$

Nous générons à nouveaux des samples suivant cette loi a posteriori via l'algorithme de Metropolis-Hastings. Les résultats sont présentés à la FIGURE 4.5. Contrairement au cas où T est fixé, ici les variables x et y semblent être moins corrélées.

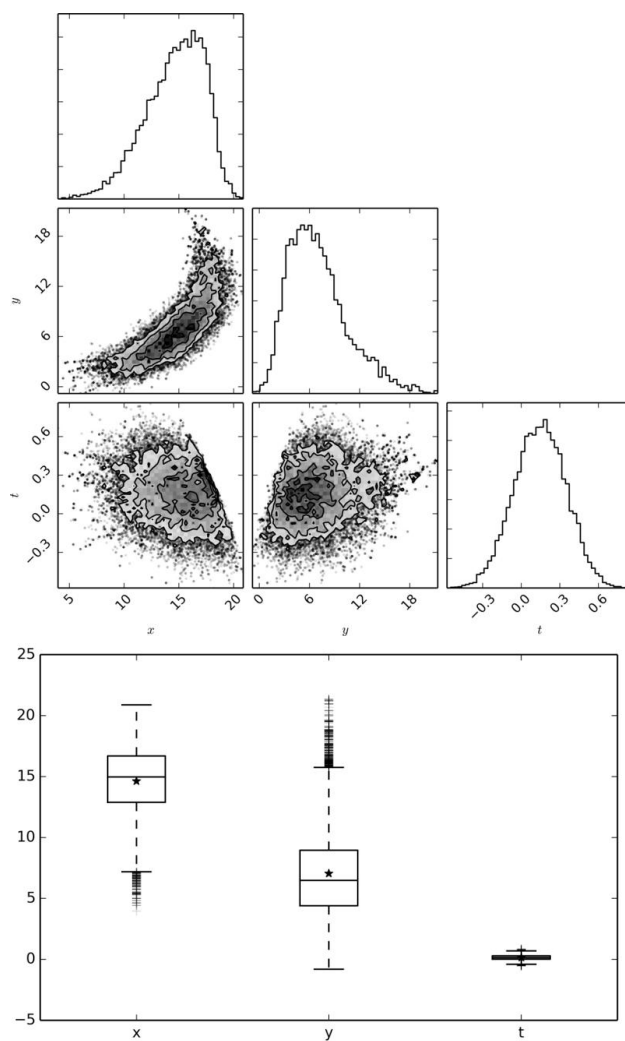


FIGURE 4.5 – Résultats de la détermination de l'endroit d'un événement sismique via inférence bayésienne, où l'on suppose que le moment du séisme est inconnu. Les graphes sont identiques à ceux de la FIGURE 4.4, avec le paramètre t supplémentaire.

Chapitre 5

Existence de l'énergie noire

Dans cette partie, nous estimons les paramètres cosmologiques, par la méthode de l'inférence bayésienne, grâce à des données observées de supernovae. Nous faisons deux tests : le premier en faisant l'hypothèse d'un univers plat, et l'autre en laissant le paramètre de courbure varier entre -1 et 1 .

5.1 Modélisation

Nous considérons un univers Λ CDM, pour rappel [3] les paramètres de densités sont $(\Omega_m, \Omega_\Lambda, \Omega_k, \Omega_r)$; ceux-ci sont liés respectivement à la matière (baryonique et noire), à la constante cosmologique, à la courbure et au fluide de radiation.

L'objectif de ce chapitre va être de déterminer les paramètres cosmologiques, sur bases d'observations réalisées sur les supernovae de type I. Nous possédons¹ comme données le module de distance de plusieurs supernova (obtenu en calculant la distance lumineuse [8]), ces données sont sous la forme (z_i, μ_i, σ_i) , où z_i est le redshift à laquelle la supernova i a été observée, μ_i est le module de distance et σ_i l'erreur de mesure de ce module, suivant un gaussienne. Il est possible de déterminer un modèle, dépendant des constantes cosmologiques et du paramètre de hubble, reliant la distance lumineuse au redshift.

Dans la métrique FLRW en coordonnées (t, χ, Ω, ϕ) , la distance luminosité s'exprime comme [3] :

$$d_L(z) = (1 + z)a_0 S_k(\chi_e(z)).$$

Où [9]

$$S_k = \begin{cases} \sin(\chi) & \text{si } \Omega_k < 0 \\ \chi & \text{si } \Omega_k = 0 \\ \sinh(\chi) & \text{si } \Omega_k > 0 \end{cases}$$

et la coordonnée comobile vaut [3] :

$$\chi_e(z) = \frac{1}{a_0} \int_0^z \frac{dz}{H(z)}$$

Le paramètre de Hubble exprimé en fonction du redshift est sous une forme

1. Fournie par mon promoteur A. Füzfa

analytique [3] :

$$H^2(z) = H_0^2(\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda + \Omega_r(1+z)^4).$$

Les données sont sous la forme de module de distance, dont la formule est [11] :

$$\mu = m - M = 5 \log_{10} \frac{d_L}{Mpc} + 25.$$

Cette dernière équation correspond au modèle, dont nous devons estimer les paramètres. Nous négligerons Ω_r [3]. Dans un premier temps, nous supposons un univers plat avec $\Omega_k = 0$. Les paramètres indépendants du modèle seront donc : (Ω_m, H_0) , et $\Omega_\Lambda = 1 - \Omega_m$. Dans un deuxième temps, nous ne feront plus l'hypothèse d'un univers plat. Nous aurons alors trois paramètres indépendants $(\Omega_m, \Omega_\Lambda, H_0)$, et $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$.

5.2 Univers plat

Comme $\Omega_k = 0$, les paramètres indépendants du modèle sont : (Ω_m, H_0) , avec $\Omega_\Lambda = 1 - \Omega_m$. Le paramètre de Hubble étant à une échelle trop petite, nous préférons utiliser h tel que $H = \frac{10^5}{Mpc} h$. L'équation du module de distance devient dès lors :

$$\mu = m - M = 5 \log_{10} \frac{d_L}{1e5} + 15.$$

Nous choisirons les prior suivantes :

$$\begin{aligned} \Omega_m &\sim U([0, 1]) \\ h &\sim U([0, +\infty[) \end{aligned}$$

Les erreurs de mesures suivent une loi normale de moyenne nulle et d'écart-types σ_i^2 données. La likelihood vaut donc :

$$L(\Omega_m, h) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-0.5 \sum_{i=1}^n \left(\frac{\mu_i - \mu_{th}(z_i, \Omega_m, h)}{\sigma_i} \right)^2 \right]$$

Les résultats sont illustrés aux FIGURES 5.1 et 5.2. On voit que la likelihood

est maximisée pour une proportion d'énergie noire supérieure à 70%. Sur le graphe de la FIGURE 5.3, nous avons comparé les observations et les données prédites par le modèle. On observe que le modèle prédit mieux les données avec les paramètres obtenus par inférence bayésienne sur base des données des supernovae, qu'en considérant 100% d'énergie noire ou 0%.

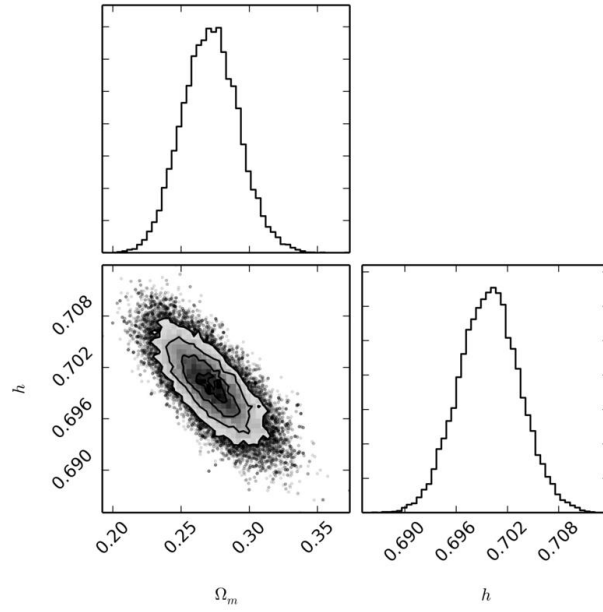


FIGURE 5.1 – Résultat de l'inférence bayésienne pour estimer la Ω_m et h dans le cas d'un univers plat.

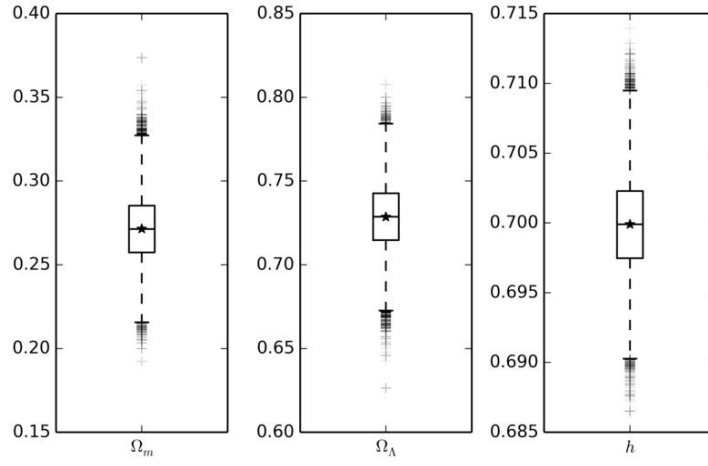


FIGURE 5.2 – Boîte à moustaches des différents samples Ω_m et h obtenus par inférence bayésienne dans le cas d'un univers plat.

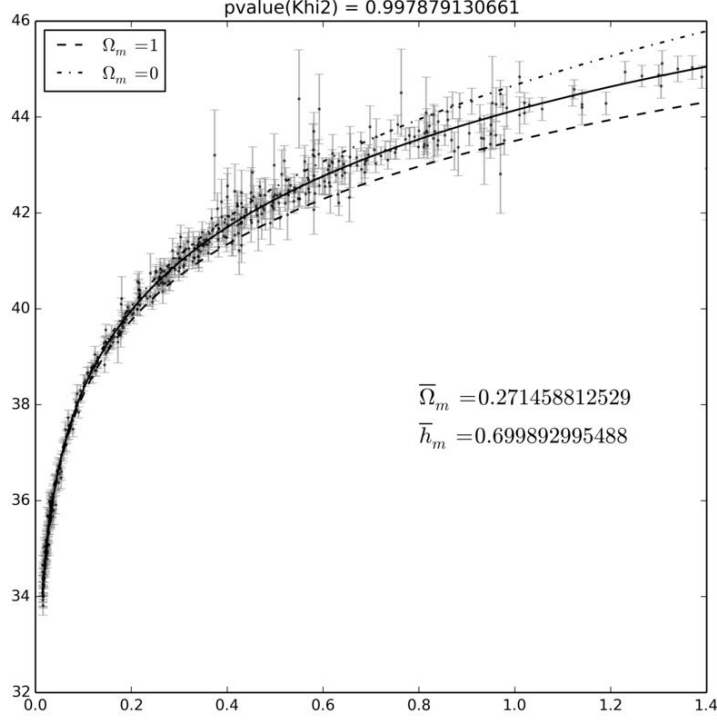


FIGURE 5.3 – Comparaison des valeurs moyennes de Ω_m et h obtenues par inférence bayésienne dans le cas d'un univers plat, avec les données des supernovae. Les simulations pour un univers avec 0% de matière noire et 100% de matière noire sont aussi représentées.

5.3 Univers sphérique, plat ou hyperbolique

Nous considérons Ω_k comme un paramètre du modèle à déterminer. Nous définissons la prior sur Ω_k comme la loi uniforme sur l'intervalle $[-1, 1]$. La likelihood vaut, cette fois :

$$L(\Omega_m, \Omega_\Lambda, h) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-0.5 \sum_{i=1}^n \left(\frac{\mu_i - \mu_{th}(z_i, \Omega_m, \Omega_\Lambda, h)}{\sigma_i} \right)^2 \right].$$

Les résultats sont illustrés aux FIGURES 5.4 et 5.5. On remarque que le paramètre de courbure est légèrement négatif (semble suivre une loi normale de moyenne -0.04447169). Ce qui ne correspond pas aux estimations ac-

tuelles [3]. Cela peut-être dû au fait que les supernovas sont observées à des redshift faibles (< 1.5). Notons que le test de Khi2 est similaire au point précédant. Lorsque l'on veut comparer deux modèles d'inférence bayésienne, on peut utiliser [15] la « Bayesian evidence », ou « Bayes factors », il s'agit du résultat d'un quotient qui, en fonction du résultat, peut privilégier un modèle plutôt que l'autre.

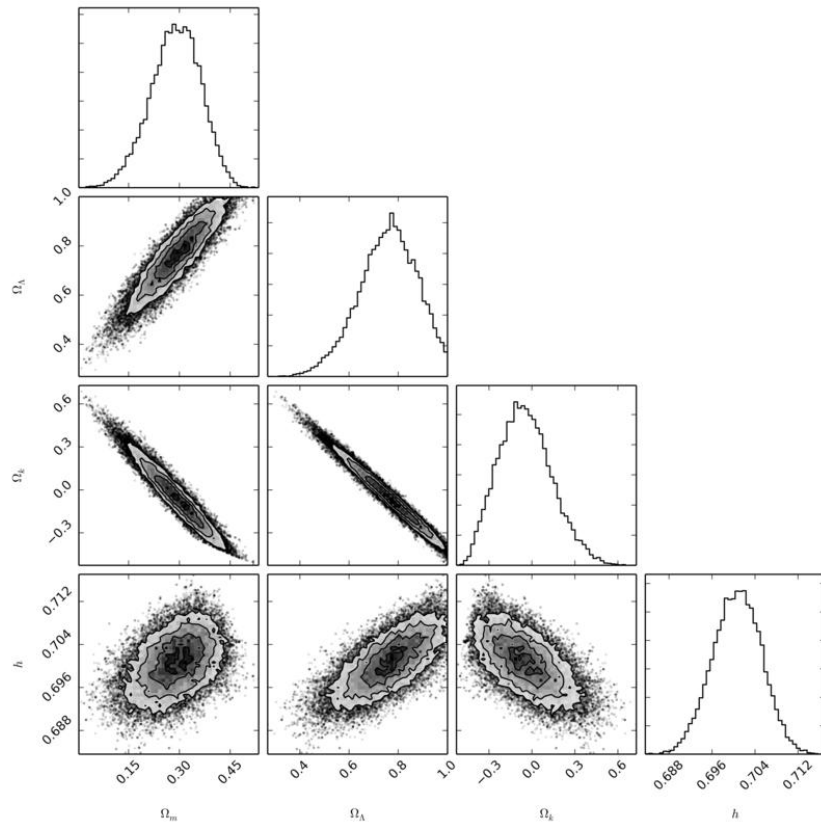


FIGURE 5.4 – Résultat de l'inférence bayésienne pour estimer $(\Omega_m, \Omega_\Lambda, h)$ où le paramètre de courbure Ω_k varie entre $[-1, 1]$

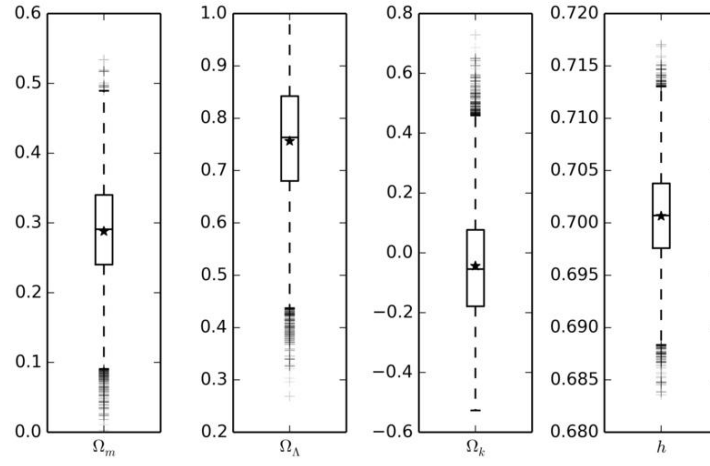


FIGURE 5.5 – Boîte à moustaches des différents samples $(\Omega_m, \Omega_\Lambda, h)$ obtenus par inférence bayésienne

Chapitre 6

Conclusions et perspectives

Sur base de la théorie de la probabilité définie comme mesure, nous avons décrit la méthode d'inférence bayésienne. Nous avons vu comment décrire de manière générale un système physique. L'inférence bayésienne est basée sur le calcul de la fonction a posteriori, définie, grâce au théorème de Bayes, par une likelihood et une prior. Une likelihood est influencée par la densité des erreurs de mesures et du modèle. Nous avons donc déterminé l'expression de ces likelihood dans le cas d'erreurs gaussiennes.

Les fonctions a posteriori obtenues n'étant pas forcément analytiques, et pouvant être des densités de probabilité de plusieurs variables, nous avons dû introduire des méthodes non classiques permettant d'engendrer des échantillons suivant n'importe quelle fonction. C'est là que l'on a alors présenté l'algorithme de Metropolis-Hastings. Celui-ci se base sur les chaînes de Markov, que nous avons également détaillées.

Nous avons dès lors toutes les notions nécessaires pour appliquer l'inférence bayésienne à des exemples concrets. On a d'abord résolu un problème de régression linéaire simple. On a vu que l'inférence ne donnait pas une solution unique, mais un ensemble de droites plus ou moins probables. Ensuite, nous l'avons appliquée à un problème plus intéressant qui consistait à détecter le centre d'un événement sismique. Nous avons d'abord fait l'hypothèse que le

moment où le séisme a eu lieu était connu, puis en l'ajoutant comme paramètre inconnu du modèle.

Lors de l'estimation des paramètres cosmologiques utilisant les données des supernovas, nous constatons que le paramètre de courbure est légèrement négatif, ce qui diffère des estimations actuelles. Nous avons également estimé les paramètres en fixant ce paramètre de courbure à zéro. Les tests χ^2 effectués donnent des résultats identiques. On a alors cité la notion de « Bayes factor », qui peut permettre de priviliger un modèle plutôt qu'un autre. La raison du paramètre de courbure négatif peut être que les supernovas sont observées à des redshift faibles (< 1.4), nous pourrions peut-être utiliser[3] les observations basées les anisotropies du fond diffus cosmologique.

Au niveau des méthodes de Monte-Carlo, il serait intéressant de réaliser des tests de convergences : en effet, une chaîne de markov converge théoriquement, en un nombre d'itération infini, vers sa distribution stationnaire. On pourrait alors avoir un ordre d'idée sur le nombre minimal de samples à générer par l'algorithme. Nous n'avons parler que de l'algorithme de MH, on aurait pu aussi citer Gibbs. Le cours de P. Protopapas [13] est d'ailleurs une source très détaillée sur le sujet.

Bibliographie

- [1] A. Hardy, *Probabilité*, Université de Namur, 2013.
- [2] J. Barthelemy, J. Boreux, *Générateurs de Nombres Pseudo-Aléatoires & Mapping d'Accélérateurs de Particules*, Université de Namur, 2012.
- [3] J.-P. Bruneton, *Gravitation relativiste et cosmologie*, Université de Namur, 2012.
- [4] S. Chakraborty, *Some Applications of Dirac's Delta Function in Statistics for More Than One Random Variable*, University of Texas Pan American, http://www.pvamu.edu/Include/Math/AAM/Vol13_No1/ChakrabortyMAA-R14-071406Final_4_6-12-08.pdf. [En ligne, dernière consultation le 15 août 2014]
- [5] W.R. Gilks, S. Richardson et D.J. Spiegelhalter, *Markov chain monte carlo in practice*, 1996
- [6] D. Foreman, D.-W. Hogg, D. Lang et J. Goodman, *emcee : The MCMC Hammer*, <http://arxiv.org/pdf/1202.3665v4.pdf>, 2013. [En ligne ; dernière consultation le 1 août 2014]
- [7] O. Garet, *Intération et Probabilités*, Université de Lorraine, 2013, <http://iecl.univ-lorraine.fr/~Olivier.Garet/cours/ip/ip-poly.pdf>. [En ligne, dernière consultation le 15 août 2014]
- [8] D. Guide, *Détermination des paramètres cosmologiques à l'aide de supernovæ de type Ia à grands décalages vers le rouge*, Université de Paris VI, Thèse de doctorat, http://lpnhe.in2p3.fr/doctorants/uploads/theses/these_11-10-05.pdf, 2005. [En ligne ; dernière consultation le 15 août 2014]

- [9] M.-P. Hobson, G.-P. Efstathiou et A.-N. Lasenby, *General Relativity. An Introduction for Physicists.*, 2006.
- [10] L. Koot, *Etude de la structure interne de la Terre à partir des observations de la précession et des nutations*, Thèse, Université Catholique de Louvain, 2009.
- [11] V. Irsic, *Supernova type Ia*, <http://fiz.fmf.uni-lj.si/~vid/opa/exercise-05.pdf>, 2013. [En ligne; dernière consultation le 15 août 2013]
- [12] W. Krauth, *Algorithms and Computations*, Oxford University Press, 2006.
- [13] P. Protopapas, *Advanced Scientific Computing : Stochastic Optimization Methods. Monte Carlo Methods for Inference and Data Analysis*, <http://iacs-courses.seas.harvard.edu/courses/am207/index.html>, 2014. [En ligne, dernière consultation le 15 août 2014]
- [14] A. Rivoldini, *The interior structure of terrestrial planets. An application to Mars*, 2009.
- [15] R. E. Kass et A. E. Raftery, *Bayes Factors*, Journal of the American Statistical Association, <https://www.andrew.cmu.edu/user/kk3n/simplicity/KassRaftery1995.pdf>, 1995. [En ligne; dernière consultation le 15 août 2014]
- [16] M. West, *Likelihood and Reference Bayesian Analysis*, 1999. [En ligne; dernière consultation le 15 août 2014]
- [17] Wikipedia, *Probabilité conditionnelle*, https://fr.wikipedia.org/wiki/Probabilité_conditionnelle#Densit.C3.A9_conditionnelle, 2013. [En ligne; dernière consultation le 10 mai 2013]
- [18] Wikipedia, *Suite arithmético-géométrique*, https://fr.wikipedia.org/wiki/Suite_arithmético-géométrique. [En ligne; dernière consultation le le 10 mai 2013]
- [19] Wikipedia, *Système physique*, https://fr.wikipedia.org/wiki/Système_physique, 2013. [En ligne; dernière consultation le 10 mai 2013]

Annexe A

Calcul de la convergence des suites

Nous déterminons ici la convergence des suites rencontrées en 3.4.1.2.

Notons x_n la première composante de u_n , et y_n sa deuxième composante, on en déduit les deux suites :

$$\begin{cases} x_0 = y_0 = 0.5 \\ x_n = \frac{1}{2}x_{n-1} + \frac{1}{10}y_{n-1} \\ y_n = \frac{1}{2}x_{n-1} + \frac{9}{10}y_{n-1} \end{cases}$$

Déterminons la convergence de ces deux suites : on peut déjà remarquer que $x_n + y_n = x_{n-1} + y_{n-1} = \dots = 1$, d'où :

$$x_n = 1 - y_n \tag{A.1}$$

De plus, on a, par définition de x_n , on a :

$$x_n = y_n + \frac{8}{10}y_{n-1} \tag{A.2}$$

En égalisant x_n des équations (A.1) et (A.2), on obtient les égalités :

$$1 - y_n = y_n - \frac{8}{10}y_{n-1}$$

$$\begin{aligned}
&\Leftrightarrow 1 = 2y_n - \frac{8}{10}y_{n-1} \\
&\Leftrightarrow y_n = \frac{2}{5}y_{n-1} + \frac{1}{2}
\end{aligned} \tag{A.3}$$

Cette dernière équation (A.3) décrit une suite arithmético-géométrique dont la résolution qui suit se base sur l'article de wikipedia sur ce type de suites [18]. Pour étudier la convergence de cette suite et se ramener à une définition de suite récurrente, on pose $w_n = v_n - v_{n-1}$. On remarque que w_n définit une suite géométrique de raison $\frac{2}{5}$:

$$\begin{aligned}
w_{n+1} &= y_{n+1} - y_n \\
&= \left(\frac{2}{5}y_n + \frac{1}{2}\right) - \left(\frac{2}{5}y_{n-1} + \frac{1}{2}\right) \\
&= \frac{2}{5}(v_n - v_{n-1}) \\
&= \frac{2}{5}w_n
\end{aligned}$$

Le lien entre la suite w_n et y_n est déterminée en calculant la somme des n premiers terme de w_n :

$$\begin{aligned}
\sum_{i=1}^n w_i &= w_1 + \dots + w_n \\
&= (v_1 - v_0) + (v_2 - v_1) + \dots + (v_{n-1} - v_{n-2}) + (v_n - v_{n-1}) \\
&= v_n - v_0
\end{aligned} \tag{A.4}$$

Comme w_n est une suite géométrique de facteur $\frac{2}{5}$, on a :

$$\begin{aligned}
\sum_{i=1}^n w_i &= \frac{w_1 - \frac{2}{5}w_n}{1 - \frac{2}{5}} \\
&= \frac{w_1 - \left(\frac{2}{5}\right)^n w_1}{1 - \frac{2}{5}} \\
&= w_1 \frac{1 - \left(\frac{2}{5}\right)^n}{\frac{3}{5}}
\end{aligned}$$

Comme $w_1 = y_1 - y_0 = \frac{2}{5}v_0 = \frac{1}{5}$, on a :

$$\sum_{i=1}^n w_i = \frac{1 - \left(\frac{2}{5}\right)^n}{3} \quad (\text{A.5})$$

Par les équations (A.4) et (A.5), on a :

$$v_n = \frac{5}{6} - \frac{\left(\frac{2}{5}\right)^n}{3}$$

On obtient donc que $\lim_{n \rightarrow \infty} y_n = \frac{5}{6}$, et, par l'équation (A.1) : $\lim_{n \rightarrow \infty} x_n = 1 - \frac{5}{6} = \frac{1}{6}$.